"Polarising Perspectives: Unveiling the Exotic in LHC Hadronic Jets"

Report Master Project Physics and Astronomy, size 60 EC,
conducted between 13-11–2023 and 16-10–2024.

Dorian Sloot

Abstract

Classifying signal versus background events and differentiating vector boson polarisation states in the fully hadronic VVjj channel may uncover evidence of new exotic particles and provide deeper insights into the dynamics of particle interactions. Through detailed feature analyses and advanced machine learning techniques, key jet features such as the DisCo score, jet mass, number of tracks, and D2 score are identified as significant enhancers of signal vs. background discrimination. For polarisation states, jet substructure variables, including $Z^{\text{cut}}_{1,2}$, $\text{Split}_{1,2}$, Angularity, and KtDR, prove most effective. Robust classification models, specifically boosted decision trees and deep neural networks, achieve receiver operating characteristic (ROC) area under the curve (AUC) scores of 0.93 and 0.90, respectively, in the signal vs. background classification. However, polarisation classification remains challenging, with ROC AUC scores around 0.7, highlighting the need for innovative techniques such as decay angle regression. Preliminary results suggest that this method may improve experimental calibration, though challenges persist with lower $|\cos\theta|$ predictions. Overall, the findings lay a solid foundation for future advancements in polarisation tagging by integrating machine learning with a thorough and comprehensive feature analysis.

STUDENTNUMBER	12733822
RESEARCHGROUP	Nikhef: ATLAS
DAILY SUPERVISOR	Dylan van Arneman, MSc
SUPERVISOR	Dr. Flavia de Almeida Dias
EXAMINER	Prof. Dr. Wouter Verkerke
STUDY	Physics & Astronomy (GRAPPA)

Vrije Universiteit Amsterdam | Universiteit van Amsterdam

Popular scientific summary

Humans have always been driven by curiosity, constantly striving to understand the world around us. One fundamental question that has persisted through the ages is where we come from and what everything is made of. From ancient ideas about elements like water, earth, fire, and air, the journey has led to the modern field of particle physics, where scientists study the smallest building blocks of the universe and the forces governing their interactions. At the heart of this field lies the remarkable Standard Model, a theoretical framework that describes particles such as quarks and leptons, which compose matter, and the forces like electromagnetism and the weak force, all in a concise formula that fits on a coffee mug.

Despite being one of the most thoroughly tested theories in physics, the Standard Model doesn't tell the complete story. It doesn't account for gravity and fails to explain the mysterious extra mass observed in the universe, known as dark matter. Therefore, scientists are searching for innovative ways to deepen understanding and extend beyond the Standard Model.

One exciting frontier in this quest is the study of polarisation, particularly that of vector bosons. Polarisation refers to the orientation of a particle's spin, which influences the angular distribution of its decay products. By analysing these patterns, specific polarisations can be filtered out much like polarised sunglasses filter light. This could potentially uncover evidence of new particles or offer deeper insights into the fundamental dynamics of particle interactions. It's like putting on "polarisation sunglasses" to reveal new aspects of nature.

Until now, much of the focus has been on leptonic decays, which involve particles like electrons and muons. However, due to the low occurrence of these events, they haven't yielded significant breakthroughs. This project takes a bold step by investigating hadronic decays, where quarks produce particles like protons and neutrons. These decays occur much more frequently, helping to overcome the issue of low statistics, but they come with their own set of challenges. As quarks turn into hadrons, the quarks can no longer be observed directly. Instead, a cascade of particles known as hadron showers appears. In the high-energy "boosted" regime, where everything happens faster and with more energy, these showers become indistinguishable, appearing as a single, dense jet of particles. Extracting useful information about the original quarks from this jet is a significant challenge.

To tackle this challenge, cutting-edge machine learning techniques are employed, specifically boosted decision trees and deep neural networks. These powerful algorithms process vast amounts of simulated data, uncovering hidden patterns that might not be obvious at first glance. A thorough feature analysis is first conducted to identify the most relevant pieces of information to feed into the models. Once the best features are selected, the models are trained, optimised, and fine-tuned to extract as much insight as possible.

While progress has been made, polarisation tagging remains a complex challenge. Current models are still refining their ability to distinguish between longitudinal and transverse polarisation states. Larger, more advanced algorithms and additional features may be required to fully solve this. For instance, this thesis makes early attempts at incorporating decay angles, a feature closely tied to polarisation, as it describes the angular distribution of decay products.

The insights gained from this work bring science closer to a deeper understanding of polarisation. This field is still in its early stages, and much more work is needed. However, the potential is enormous. Polarisation tagging could open new doors in particle physics, helping to unravel some of the universe's most profound mysteries. This ongoing adventure of discovery reminds us that even the tiniest particles hold the secrets of the universe, waiting to be uncovered.

Contents

1 Introduction

Since the dawn of humanity, there has been a continuous effort to understand how the world works. Initially, ancient civilisations believed that everything was composed of four basic elements: water, earth, fire and air. However, as our understanding deepened, it became evident that the world was composed of more intricate building blocks. This revelation led to the discovery of over a hundred chemical elements throughout the 19^{th} century, leading to the modern periodic table known today [1].

The discovery of the electron in 1897 revolutionised our perspective, revealing that the essence of nature was composed of particles even smaller than previously imagined. In the years that followed, a variety of subatomic particles has been discovered and categorised within the Standard Model (Section 2.1). This theoretical framework effectively describes the properties and interactions of the matter surrounding us, condensed into a single mathematical equation. This crucial moment marked the beginning of modern particle physics, paving the way for an exploration of the true building blocks of nature [2].

1.1 Particle Physics

Particle physics, also known as high-energy physics, explores realms beyond our everyday experience. Many of the particles studied cannot be directly observed outside atomic nuclei under normal conditions. Instead, they are created and detected within controlled laboratories known as particle accelerators. By accelerating particles to incredibly high speeds and colliding them, physicists analyse the resulting debris to unravel the fundamental nature of matter and the forces governing subatomic particles, offering a unique glimpse into the underlying structure of the universe.

1.2 CERN

At the forefront of particle physics research stands the European Organisation for Nuclear Research, CERN, regarded as the largest and one of most respected research centres in the world. CERN conducts high-energy physics research using particle accelerators and other infrastructures, with the greatest asset being the Large Hadron Collider (LHC). The LHC is the world's largest and highest-energy particle collider and is home to a wide range of experiments run by collaborations of scientists from institutes worldwide. The four biggest and most prominent experiments are ATLAS, CMS, ALICE, and LHCb. Each one of these experiments is unique and characterised by its detectors. ALICE and LHCb are equipped with specialised detectors tailored to focus on specific phenomena, while ATLAS and CMS utilise general-purpose detectors to explore the widest spectrum of physics possible.

1.3 The ATLAS experiment

ATLAS is designed to pick up a wide variety of signals to ensure it can detect and study any new physical processes or particles that might appear. It is built to identify different features of these particles, such as their energy, electrical charge, lifetime, mass, momentum and spin. To achieve this, the detector is composed of multiple specialised layers (Fig. 1), each serving a distinct purpose in the particle detection process.

At the core of the detector lies the Inner Tracker, which is responsible for tracking the paths of charged particles as they pass through the detector. This layer is crucial for determining the momentum of particles by measuring the curvature of their tracks in the presence of a magnetic field, which bends the paths of charged particles.

Surrounding the inner tracker are two calorimeter layers: the Electromagnetic and the Hadronic Calorimeter. The electromagnetic calorimeter measures the energy of particles that interact via electromagnetic forces, such as electrons and photons. On the other hand, the hadronic calorimeter measures the energy of hadrons, such as protons and neutrons, which interact via the strong force. These layers are essential for reconstructing the energy and type of particles produced in the collisions.

On the outermost part of the detector is the Muon Spectrometer, designed specifically to detect and measure the properties of muons, which are a heavier generation of electrons. Muons are particularly significant because they can penetrate the inner layers of the detector due to their high mass and relatively weak interaction with matter, making the outer layer ideal for their identification and measurement.

Together, these layers work in unison to provide a comprehensive picture of the particles produced in collisions. By harnessing the unprecedented energy provided by the LHC, ATLAS is able to explore the fundamental properties of matter with greater precision than ever before. This powerful combination allows ATLAS to uncover new physics phenomena that were previously beyond reach.

The experiment's capabilities were showcased in 2012 when it played a crucial role in the discovery of the Higgs boson, a particle that had long evaded detection by scientists. This breakthrough highlighted ATLAS's significance in deepening our understanding of particle physics. Moreover, the experiment is set to delve even deeper into the mysteries of the universe. Where for familiar processes, ATLAS works on refining measurements of known particles or obtaining numerical validations of the Standard Model. On the other hand, the detection of previously unseen processes could reveal new particles or offer supporting evidence for theories beyond the Standard Model. Thus, ATLAS's research has the potential to both strengthen current understanding of particle physics and open up new avenues of exploration.

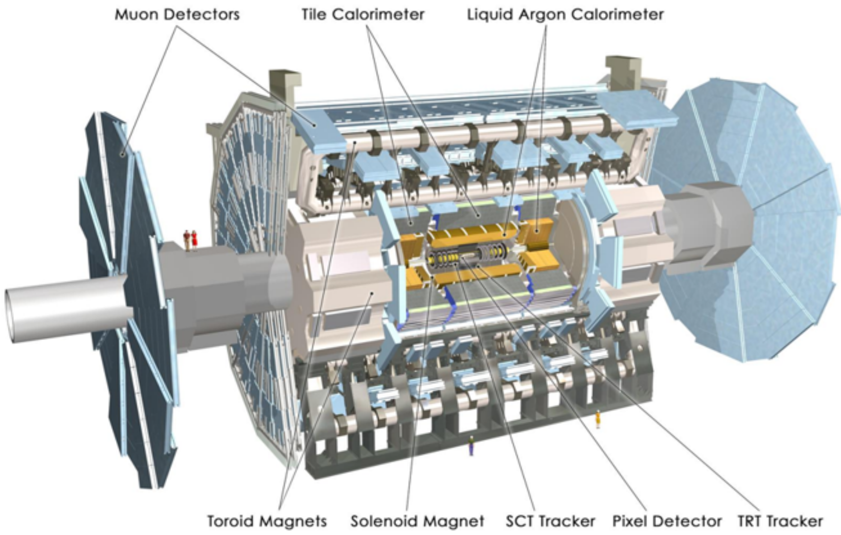


Figure 1: Schematic of the ATLAS detector, illustrating its multiple layers [3]. Each layer is designed to identify and measure different properties of particles produced in collisions. The inner detector tracks particle trajectories, the calorimeters measure energy, and the muon spectrometer detects muons with high precision. Together, these layers provide comprehensive data crucial for particle physics research.

1.4 A Polarising Search

In line with ATLAS's overarching goals to advance particle physics research, this thesis aims to contribute to the ongoing exploration of theories beyond the Standard Model. A promising avenue of investigation is the study of polarisation patterns within events, which may reveal evidence of previously undiscovered particles or provide insights into the fundamental dynamics of particle interactions. This approach offers a fresh perspective for probing the data, aiding in the search for physics beyond the Standard Model (BSM).

Specifically, this research focuses on exotic particles characterised by the production of two vector bosons from the decay of a heavy resonance. The investigation centres on the fully hadronic channel, where both vector bosons decay into pairs of quarks, resulting in hadronic jets. This strategic choice enables larger statistics and enhances sensitivity to multiple BSM theories simultaneously.

By studying exotic particles through polarisation identification, aspects of ATLAS data that have not been previously observed will be revealed. Valuable insights into electroweak symmetry breaking could be retrieved from longitudinally polarised W bosons. Additionally, high-energy diboson interactions may conceal new physics that could be uncovered by strategically analysing polarisation fractions to detect the tails of resonance peaks beyond the LHC energy reach. Consequently, this work contributes to the ongoing exploration of exotic particles and lays the groundwork for future searches using polarisation identification techniques.

2 Theory

2.1 The Standard Model

The Standard Model stands as the cornerstone of modern particle physics, representing the most successful theoretical framework for explaining the behaviour of the matter around us. At its core, the Standard Model encapsulates the fundamental building blocks of matter and the interactions between them [4], known as the force mediators. These elementary particles are categorised into two subgroups: fermions and bosons (Fig. 2).

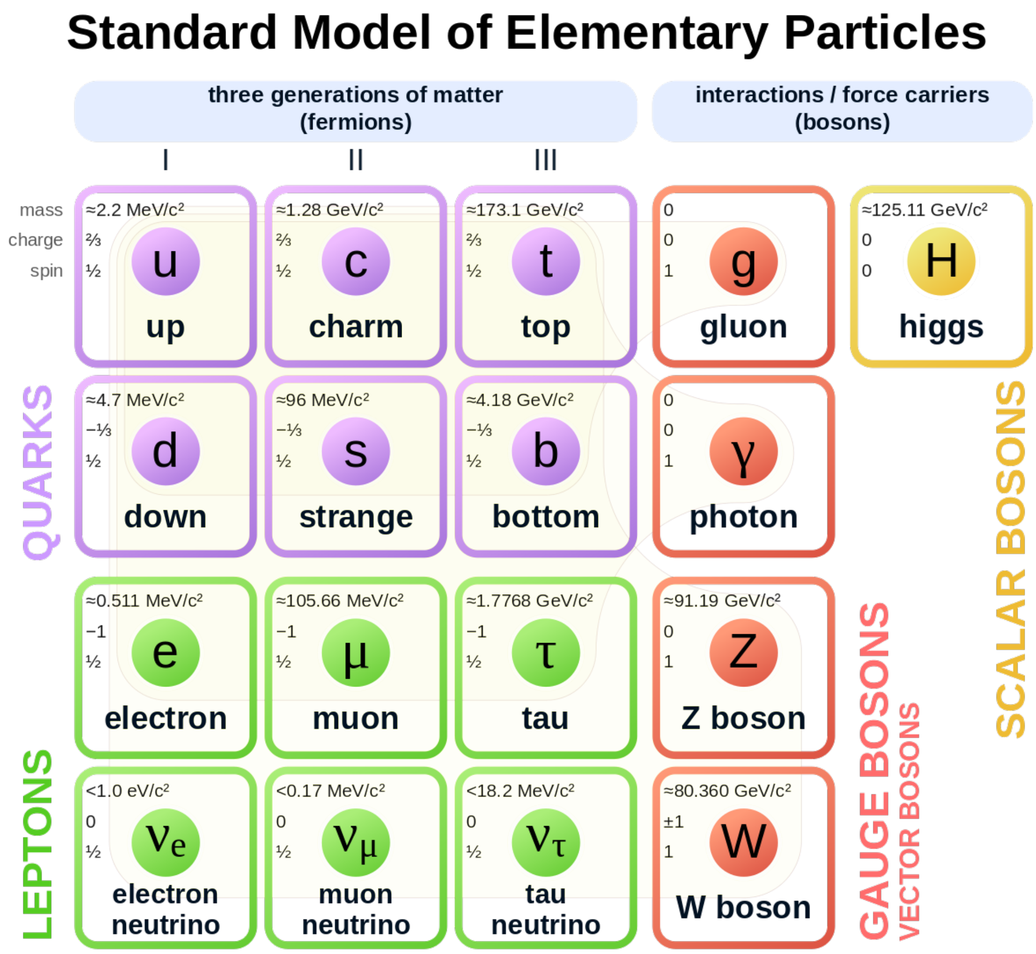


Figure 2: The Standard Model of particle physics, illustrating the elementary particles categorised into fermions and bosons [5]. Fermions include quarks and leptons, which constitute matter, while bosons serve as force carriers responsible for mediating the fundamental interactions: the photon for electromagnetism, the W and Z bosons for weak interactions, and the gluons for the strong force. The Higgs boson, which is responsible for giving mass to particles through the Higgs mechanism, is also depicted.

2.1.1 Fermions

Fermions, characterised by their spin-1/2 property, serve as the fundamental building blocks of matter. They are further divided into two subcategories, leptons and quarks, each featuring three generations with increasing masses, as shown in Figure 2.

There are six leptons, classified by their electric charge (Q), electron number (L_e), muon number (L_μ), and tau number (L_τ). Each generation consists of a charged particle (Q = -1), namely the electron, muon, and tau respectively (e^-, μ^-, τ^-), and their corresponding neutral neutrinos ($\nu_{e^-}, \nu_{\mu^-}, \nu_{\tau^-}$). The first generation contains the lightest and most stable particles, whereas the heavier and less stable ones belong to the second and third generations. As the heavier particles quickly decay to more stable forms, all stable matter is made of first generation particles. Furthermore, each lepton has its own antiparticle, possessing identical mass and spin but opposite charges (both electric and colour).

Similarly, there are six 'flavours' of quarks and antiquarks, classified by charge, strangeness (S), charm (C), beauty (B), and truth (T). Each generation comprises an up-type particle (up, charm, top) with electric charge Q = $+\frac{2}{3}$ and a down-type particle (down, strange, bottom), with Q = $-\frac{1}{3}$. Unlike leptons, quarks possess an additional charge called colour, associated with the strong force, which comes in three variations: red, green, and blue. The colour analogy comes from the fact that quarks cannot exist freely as stable particles but instead combine to form colour-neutral hadrons. These hadrons are then divided into two main categories: baryons, composed of an odd number of quarks (typically three: $q_\mathrm{r} q_\mathrm{g} q_\mathrm{b}$), and mesons, composed of an even number of quarks (typically two: $q_c \bar{q}_{\bar{c}}$).

2.1.2 Bosons

In contrast to fermions, bosons are particles with integer spin that act as the mediators of the fundamental forces. In modern particle physics, each force is described by a Quantum Field Theory (QFT) involving the exchange of a spin-1 force-carrying particle [4], known as a gauge boson. For electromagnetism, this is the theory of Quantum Electrodynamics (QED), where interactions between charged particles are mediated by the exchange of photons. Similarly, the weak force is described by Quantum Flavour Dynamics (QFD) and mediated by the charged W^+, W^- and the electrically neutral Z bosons. In practice, electromagnetism and the weak interaction are often discussed within their unified framework, the Electroweak Theory (EWT). The final fundamental force included in the Standard Model is the strong force, described by the QFT of Quantum Chromodynamics (QCD) and mediated by gluons. Like quarks, gluons carry colour charge and cannot exist as free particles. Consequently, gluons can only be detected within hadrons or in colourless combinations with other gluons, known as glueballs [2].

The final piece of the Standard Model is the spin-0 scalar boson known as the Higgs boson. This particle stands out due to its unique role in providing mass to other particles. Without the Higgs, the universe would be completely different, with all particles being massless and traveling at the speed of light. In Quantum Field Theory, the Higgs boson is seen as an excitation of the Higgs field, which differs from other fields by having a non-zero vacuum expectation value. The interaction between initially massless particles and the non-zero Higgs field gives rise to their masses. The masses of the Higgs boson, along with the $W^{\pm}$ and Z bosons, are around 100 GeV and collectively define the electroweak scale [4]. This relationship is not coincidental because, in the Standard Model, the masses of weak gauge bosons are intricately linked to the Higgs mechanism.

2.2 Beyond the Standard Model

Despite its numerous successes, the Standard Model has significant shortcomings, making it an incomplete description of nature. Consequently, the search for physics beyond the Standard Model is essential. The primary shortcomings of the Standard Model and the main reasons for seeking new physics are outlined below.

First of all, while the model is theoretically expected to describe everything, it currently only accounts for three out of the four known fundamental forces. This discrepancy arises from the absence of a comprehensive quantum description of gravity [6]. However, given that Einstein's theory of general relativity describes gravity as the weakest among the fundamental forces, its coupling strength to elementary particles can be neglected. Nonetheless, it dominates at large distance scales and therefore should be included in any unified theory.

Secondly, the Standard Model cannot explain the apparent matter-antimatter asymmetry in the universe. This refers to the imbalance in the quantities of particles and antiparticles in the observable universe. While certain asymmetries can be rationalised by established Standard Model mechanisms involving charge conjugation and parity symmetry (CP) violation, the extent of these processes observed thus far fails to fully explain the observed asymmetry [7].

Moreover, apart from the observed asymmetry, a remarkable 85% of the mass in the universe is comprised of dark matter [8, 9], yet it is not described by the Standard Model. Despite lacking direct observation, their existence and properties can be inferred through indirect measurements. Various hypotheses have been proposed regarding the nature of the particles composing dark matter, but none have been empirically confirmed. Although indications suggest dark matter may consist of particles, the particles in the Standard Model do not possess the right properties to form it [10].

Additionally, there are challenges in aligning the model to some experimental results. The discovery of neutrino oscillations in 1998 provided definitive proof that neutrinos have mass

[11], a phenomenon not predicted by the Standard Model. To account for this, two possible approaches can be considered, depending on the intrinsic nature of the neutrino. Neutrinos could be Dirac particles, where the neutrino and its antiparticle are distinct ($\nu \neq \bar{\nu}$), or Majorana particles, where the neutrino is its own antiparticle. The latter scenario having multiple implications, including potential explanations for the matter-antimatter asymmetry and the existence of a very heavy neutrino as predicted by the see-saw mechanism [10].

Finally, the Standard Model faces an issue known as the hierarchy problem [12]. This problem arises when there is a large discrepancy between the expected value of a parameter in a theoretical model and its measured value. In particle physics, this occurs when trying to explain the difference between the mass of the Higgs boson and the Planck mass. Extensions of the Standard Model predict quantum corrections that significantly affect the Higgs boson's mass. The effective mass of the Higgs boson is the sum of its bare mass and these quantum corrections. If the Standard Model holds up to the Planck scale, a precise fine-tuning of about 10^{17} would be required for the Higgs boson's bare mass to match the observed mass [4]. While this is not a problem in itself, such extreme fine-tuning is considered unnatural.

All these limitations of the Standard Model have driven physicists to develop new theories to extend its framework. To determine which extension of the Standard Model aligns with reality, both experiments and precise theoretical predictions are essential. However, the search for BSM signals is inherently challenging due to several factors [13]. Primarily, the nature of new physics is unknown, raising the question of how to search for something undefined. Additionally, even with a hypothetical target, the stochastic nature of particle physics means we can only predict the probabilities of specific outcomes, not the exact results of individual interactions. Furthermore, the complexity of these experiments makes it difficult to directly calculate the probability that any specific model configuration is realised in nature.

2.3 Event Signature

A common approach in particle physics is to optimise the search for certain signals from specific models, enhancing statistical sensitivity by eliminating irrelevant measurements and reducing background noise. This targeted approach allows researchers to explore the significance of specific signals in understanding particle interactions.

Within the realm of diboson analysis, one such targeted approach focuses on the production of two vector bosons from the decay of a heavy resonance [14, 15]. Here, the final state of the events is characterised by the subsequent decay of the vector bosons. Specifically, the fully hadronic channel, also known as VVjj, involves both vector bosons decaying into pairs of quarks, resulting in hadronic jets (Fig. 3). With the highest branching ratio (BR($W \to q\bar{q}$) = 68% and BR($Z \to q\bar{q}$) = 69%), this channel provides the largest statistics. However, its effectiveness is challenged by substantial background noise from the overwhelming number of QCD dijet events.

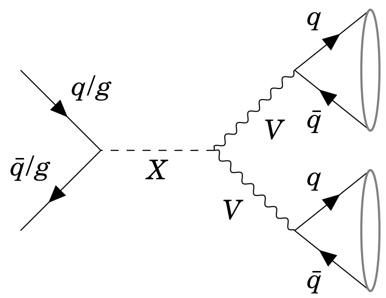


Figure 3: Feynman diagram of the fully hadronic channel from a heavy resonance decaying into two vector bosons. Here, X represents the heavy resonance, V denotes the vector bosons (either $W^{\pm}$ or Z), and q indicates the quarks. The circles connecting the quarks symbolise a Large-R jet, where the hadronic jets from the quarks are combined in the boosted regime [16].

Several BSM theories predict the existence of a new heavy particle decaying into a pair of vector bosons (WW, WZ, ZZ). Although the search strategy is entirely signature-based and not dependent on a specific theoretical model, benchmark scenarios are necessary for interpreting the results. Models that could account for this signature include the Extended Gauge Symmetry Model [17], Randall-Sundrum Model (Radion Model), Bulk Randall-Sundrum Model [12], and Heavy Vector Triplet Model [18]. However, none of these models have been conclusively confirmed or disproved. Instead, only exclusion limits have been established.

2.4 Vector Boson Polarisation

Given the lack of conclusive validation of the aforementioned models, there is a clear need for alternative methods. One approach is exploring polarisation effects. Through examination of the polarisation properties of the particles involved, a deeper understanding of the underlying physics could be gained.

A particle is referred to as polarised if its spin S has a preferred direction with respect to its momentum $\vec{p}$. This polarisation is quantified using helicity h:

$$h = \vec{S} \cdot \frac{\vec{p}}{|p|}. \tag{2.1}$$

For vector bosons, the helicity can take on one of three values: $h = \pm 1$, $h = 0$. Particles with $h = 0$ are longitudinally polarised, while those with $h = \pm 1$ are transversely polarised. Among the transverse polarisations, $h = +1$ is referred to as right handed, and $h = -1$ as left handed.

Given the conservation of spin, there is a direct relationship between the helicity of a particle and the angular distribution of its decay products [19]. Since helicity inherently involves the momentum vector $\vec{p}$, this relationship is dependent on the frame of reference.

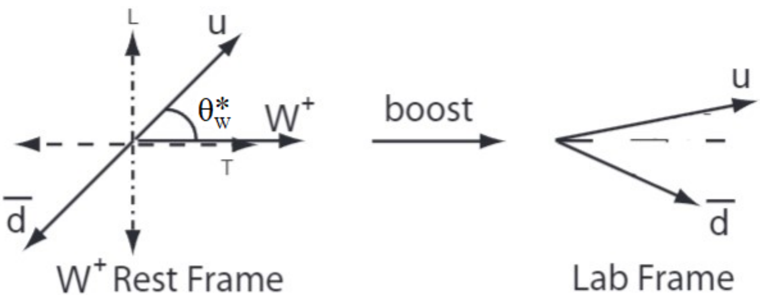


Figure 4: Schematic showing the directions of the momenta and spins of the decay products of a W_L^+ boson in both its rest frame and the lab frame [20]. On the left, the diagram highlights the polar angle $\theta_W^\star$, which represents the angle between the mother particle and its decay products in the rest frame.

Consider the decay $W^+ \to u\overline{d}$ in the rest frame of the W^+, as shown in Figure 4. The angular distribution of the decay products depends on a single degree of freedom in the decay plane, which can be chosen as the polar angle θ_V^*. This angle is defined as is the angle between the direction of the momentum of the boson and the quark's momentum in the boson's rest-frame (Fig. 5). For a given lab frame helicity, the amplitude for W^+ decay depends on $\theta_V^\star$ as follows [20]:

$$M_\pm \propto \frac{1 \mp \cos\theta_V^\star}{2}, \tag{2.2}$$

$$M_0 \propto -\frac{\sin\theta_V^\star}{\sqrt{2}}. \tag{2.3}$$

Here, the subscripts $M_{\pm,0}$ denote the helicity of the W^+. Similar expressions apply to the helicity amplitudes of W^- decays, where the polar angle is defined as the angle between the antiquark's momentum axis in the W^- rest frame and the direction of its boost.

The distributions of the $\theta_V^\star$ angles, illustrated in Figure 6, show that the decay products from longitudinally polarised vector bosons tend to decay perpendicularly to the direction of the parent bosons. In contrast, for transverse modes, the decay products align predominantly parallel or anti-parallel. This characteristic plays a crucial role in distinguishing vector boson polarisations in the lab frame.

Boosting the decay products to the lab frame results in the configuration shown in Figure 4. Due to preferential (anti-)parallel decays, decay products of transversely polarised vector bosons typically exhibit a larger energy difference between the quark and antiquark, as well as a wider opening angle in the lab frame. Meanwhile, longitudinally polarised bosons tend to distribute energy more evenly, resulting in minimal asymmetry. This energy difference serves as a lab frame observable for probing vector boson polarisation.

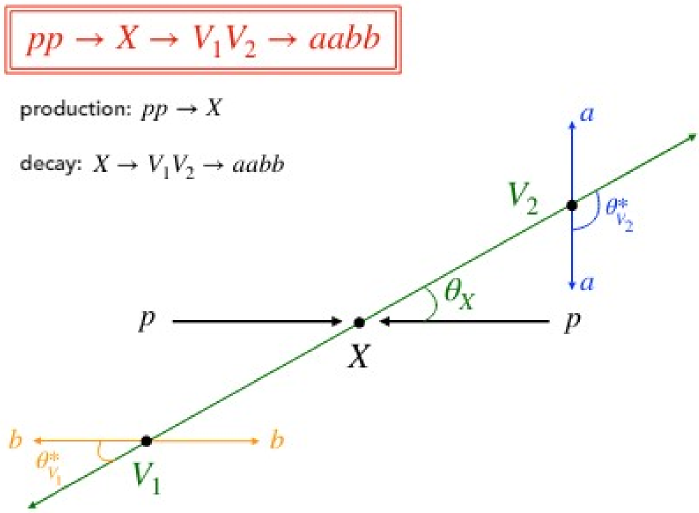


Figure 5: Schematic view of the polarisation axis in the production of a heavy resonance X from a proton-proton collision, which subsequently decays into two vector bosons [21]. These vector bosons further decay into pairs of daughter particles aa and bb. Experimentally, polarisation can be approximated as the correlation between the beam axis (green), the parent particle axis (red), and the decay product axes (yellow and blue), denoted by the angles θ_X and $\theta^\star_{V_{1,2}}$.

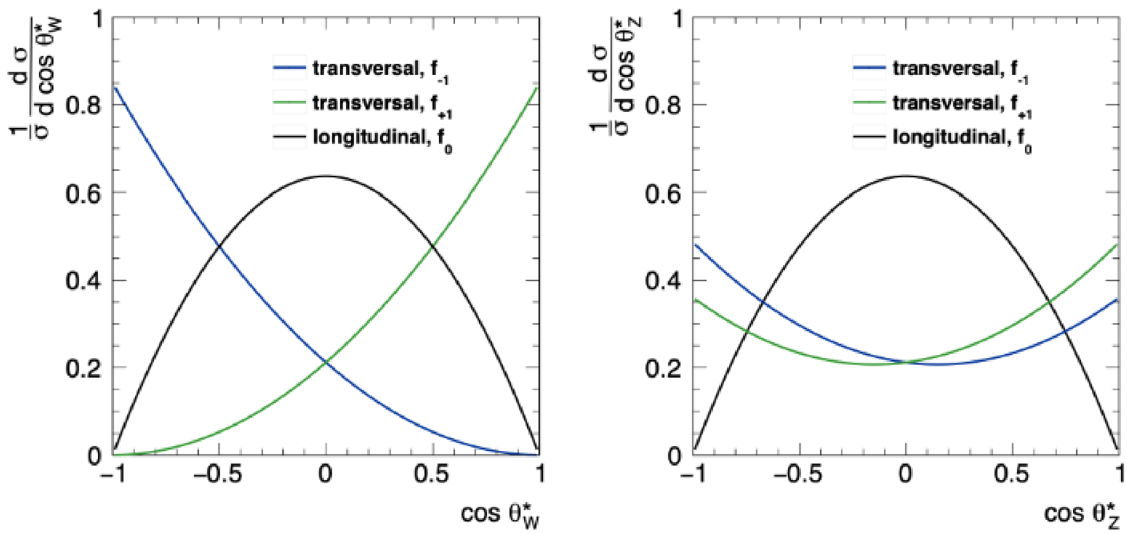


Figure 6: Distributions illustrating distinct decay patterns that distinguish vector boson polarisations in the lab frame [19], where f_h corresponds to the polarisation fractions of the helicity state h = (1, 0, -1). Longitudinally polarised vector bosons exhibit peaks at $\cos\theta^\star_{W,Z} = 0$ (corresponding to $\theta^\star_{W,Z} = \pm 90°$), indicating decay perpendicular to the parent axis. In contrast, for transversely polarised bosons aligned (anti-)parallel to the parent axis, the distribution peaks at $\cos\theta^\star_{W,Z} = \pm 1$ (corresponding to $\theta^\star_{W,Z} = 0°$ and 180°).

A direct relationship is found between $\theta_V^\star$ and $\Delta E \equiv E_q - E_{\bar{q}}$, the energy difference between the quark and the antiquark in the lab frame at parton level:

$$\cos\theta_V^\star = \frac{\Delta E}{p_V}, \tag{2.4}$$

where p_V denotes the momentum of the vector boson in the lab frame [20]. This equation connects the rest frame variable $\theta_V^\star$ to lab frame observables in a straightforward manner.

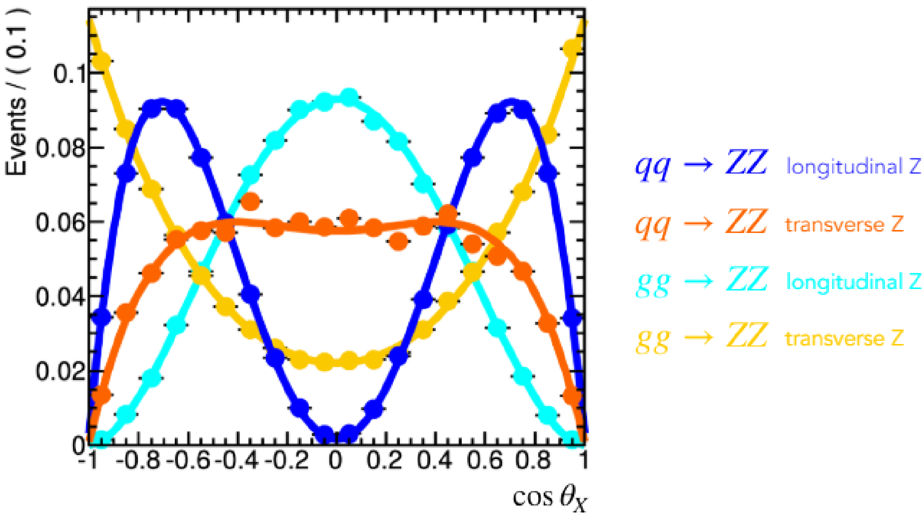


Figure 7: Graph showing the process dependence of the angular probability distribution for polarised Z bosons produced via both gluon fusion and quark-antiquark annihilation [22].

The angle θ_X between the beam axis and the vector bosons is also related to the polarisation of the vector bosons, as shown in Figure 7. However, the distribution of this angle is process dependent, changing with the polarisation of the vector bosons and the nature of the primary collision (quark-quark, gluon-gluon, or quark-gluon) as well as the decay mode of the hypothetical new particle.

In the context of the VVjj analysis, the θ_X angle can be used for final state selection, yet it requires specific interpretation. The $\theta_V^\star$ angle provides jet-level information about polarisation and has been extensively studied in the context of vector bosons decaying leptonically, where it can be precisely measured [19]. This angle can also be used for fully hadronic decays at low momentum, where the two quarks can be reconstructed as distinct small-R jets, leading to an expected decay rate distribution of:

$$\frac{1}{\sigma}\frac{d\sigma}{d\,|\cos\theta_V^\star|} = f_T\frac{3}{4}(1+|\cos\theta_V^\star|^2) + f_L\frac{3}{2}(1-|\cos\theta_V^\star|^2), \tag{2.5}$$

where f_T and f_L are the transverse and longitudinal polarisation fractions, respectively [20].

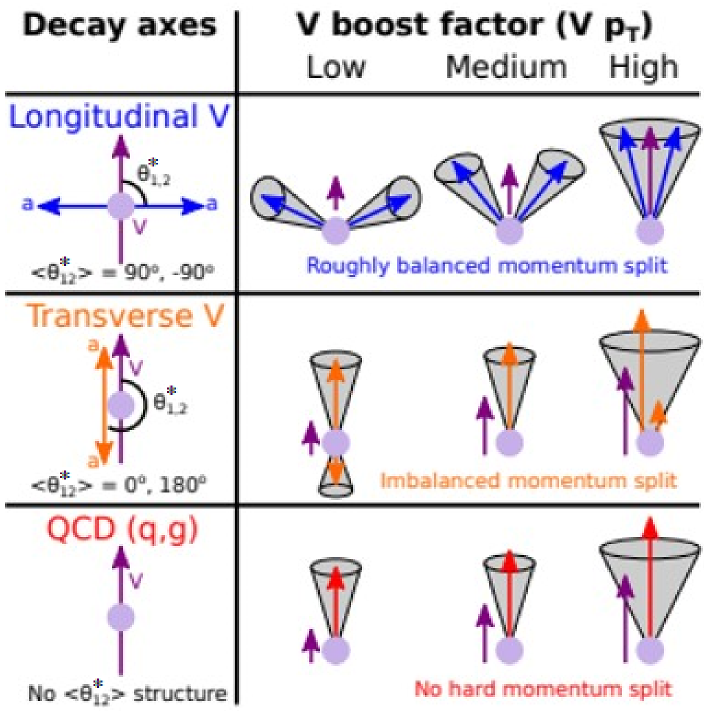


Figure 8: Schematic depicting the decay axis and boost factors of polarised vector bosons alongside the QCD background [22]. The longitudinal and transverse polarisations of the bosons significantly influence their decay angles, $\theta^{\star}_{V_{1,2}}$. These angles correlate with the differences between the boost and decay axes, impacting the substructure of the bosons. At high transverse momentum (p_T), all scenarios lead to the formation of a Large-R jet, complicating the measurement of substructure observables.

At high p_T, the decay products of boosted vector bosons are highly collimated and typically identified as a single fat or Large-R jet, as shown in Figure 8. This complicates measuring the angle between them and their Lorentz transformation back to the boson's rest-frame. To construct the angular distribution, a proxy variable for the partonic-level decay angle $|\cos\theta^{\star}_{V}|$ needs to be defined, following Equation 2.4. This involves using both the hadron-level momentum reconstruction p_V^{reco} and the subjet energy difference $|\Delta E^{\text{reco}}|$ of the Large-R jets [20].

To construct a proxy for $|\cos\theta^{\star}_{V}|$, any substructure observable that mimics the energy difference of parton-level objects can be used. Some Jet Substructure (JSS) variables, such as $\text{Split}_{1,2}$, $\tau_{2,1}$, $Z^{\text{cut}}_{1,2}$, and PtImbalance, show sensitivity to this angle and can differentiate between different polarisations [21]. However, several factors can alter the distribution of the proxy variable, including colour connections of decay products, initial state radiation (ISR), pile-up effects, detector uncertainties, and jet pruning. These factors can distort the expected parton-level distributions, affecting the accuracy of polarisation tagging. Therefore, advanced methods such as machine learning need to be explored to improve the accuracy and reliability of polarisation measurements.

2.5 Machine learning

When traditional approaches fail to solve complex problems, machine learning (ML) becomes a valuable alternative. This branch of artificial intelligence (AI) uses data and algorithms to mimic human learning processes, gradually improving its accuracy. Machine learning can quickly analyse vast amounts of data, uncovering patterns that may not be evident to human analysts. This makes it particularly valuable in fields like particle physics, where the complexity and volume of data exceed human capacity.

Machine learning applications are categorised into two primary approaches: supervised and unsupervised learning. Additionally, there are less common types such as semi-supervised and reinforcement learning, which will not be covered in this thesis. These approaches differ in how the models are trained and the type of training data that is required.

Unsupervised learning uses unlabelled datasets to train algorithms, requiring them to uncover patterns without external guidance. For instance, an algorithm in particle physics might analyse raw data from detectors at a collider experiment to identify underlying patterns in particle interactions without prior knowledge of specific particles or interactions. A well-known unsupervised learning task is clustering, illustrated in Figure 9_{a}. This method groups data points into clusters based on their inherent distribution within the dataset, making it valuable for scenarios where predefined groups are absent.

In contrast, supervised learning requires predefined groups, as algorithms are trained on labelled datasets that include tags describing each piece of data. The two most common supervised learning tasks are classification and regression. In classification, the objective is to assign a data point to a particular class, where each class represents a discrete category (Fig. 9_{b}). Examples include particle identification (e.g., determining what kind of particle it is) and signal vs. background discrimination (e.g., distinguishing signal from background events). On the other hand, regression tasks involve predicting continuous quantities associated with each data point (Fig. 9_{c}). Such tasks include energy estimation, mass reconstruction, and trajectory prediction.

Each approach possesses unique strengths, making them suitable for facing different tasks and challenges. These tasks can employ a variety of architectures and algorithms. Common examples include decision trees and deep neural networks (DNNs). Decision trees segment the data into subsets based on input feature values, creating a tree structure that enables sequential decision-making (Fig. 10). A notable variant of decision trees is the boosted decision tree (BDT). BDTs employ boosting techniques to combine multiple weak learners (simple decision trees) into a strong classifier. In boosting, each decision tree is trained sequentially, with each subsequent tree correcting the errors of its predecessors, thereby improving overall model performance and accuracy. Meanwhile, deep neural networks utilise layers of interconnected nodes (neurons) to capture intricate patterns within data (Fig. 11). This is particularly effective for tasks requiring complex pattern recognition and abstraction.

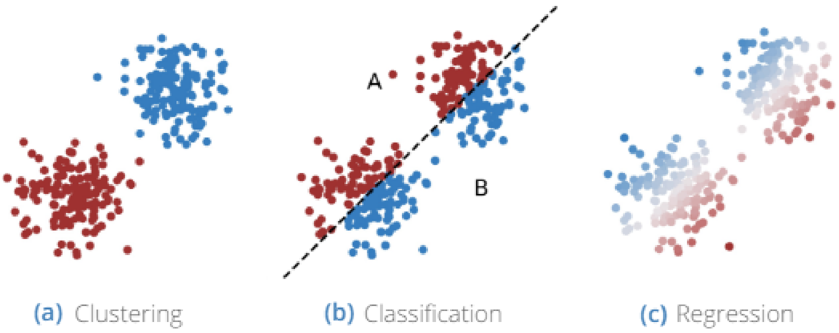


Figure 9: Common tasks in the application of machine learning [13]. (a) Clustering organises data into distinct groups based on similarities, revealing the natural data distribution without predefined categories. (b) Classification assigns data to predefined discrete categories (illustrated by red and blue dots), independent of the data distribution. (c) Regression predicts continuous outcomes (visualised as a colour gradient), such as estimating a value like momentum.

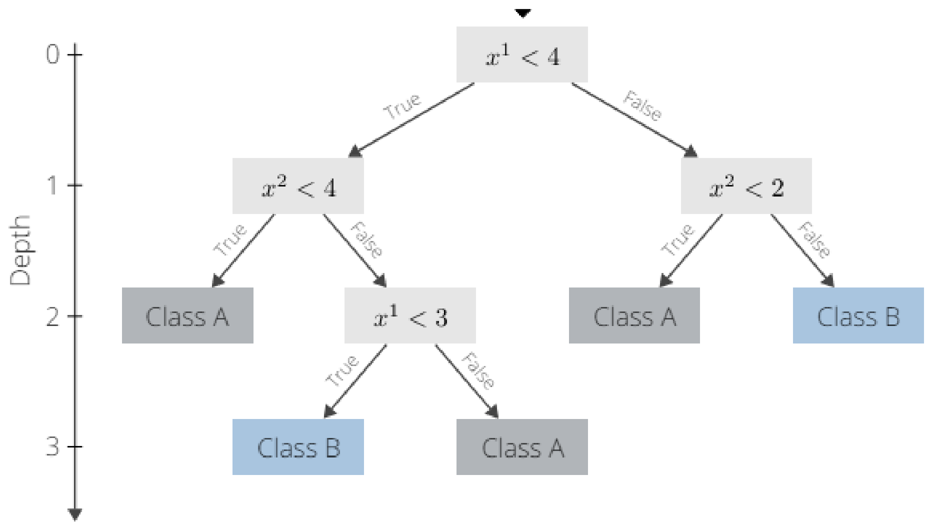


Figure 10: Schematic representation of a decision tree model for binary classification with a maximum depth of 3 [13].

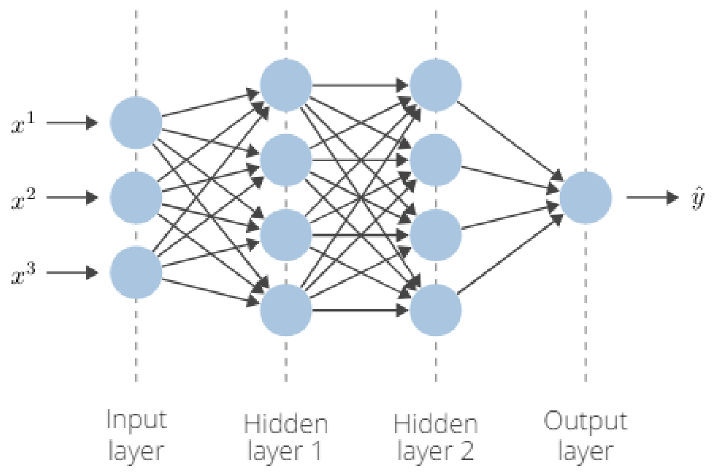


Figure 11: Schematic of a deep neural network with four layers [13]. The network consists of an input layer with three neurons, two hidden layers with four neurons each, and an output layer with one neuron for binary classification.

2.5.1 Training

Training machine learning models involves several crucial steps to ensure effective performance. The process begins with defining the task, whether it is classification, regression, or clustering, and selecting the most suitable model, such as a BDT or DNN. The architecture of the chosen model, including the number of layers and the types of activation functions (Fig. 12), is essential in determining the model's ability to capture and learn from the data. Optimising model performance also requires careful feature engineering and data tuning. This process involves selecting, transforming, and preprocessing features to ensure the model is trained on the highest quality inputs.

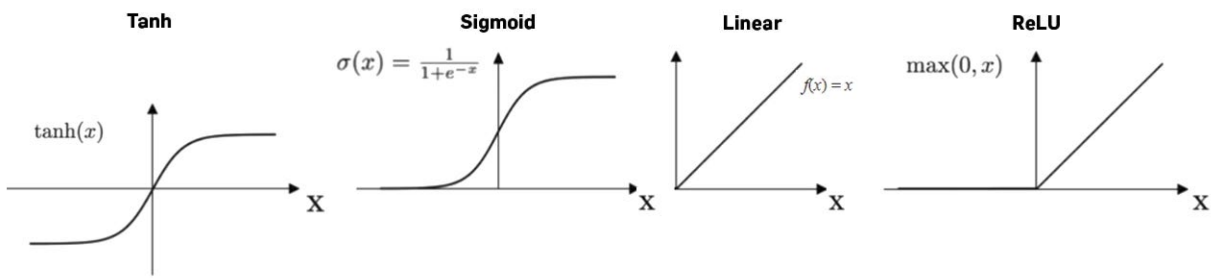


Figure 12: Common activation functions used in neural networks for classification and regression. The first function, tanh, outputs values between -1 and 1, and is applied in both tasks. The second function, sigmoid, produces outputs between 0 and 1 and is commonly used in binary classification. The linear function directly outputs the input and is typically used in regression. Lastly, ReLU (Rectified Linear Unit) returns the input if positive, or zero otherwise, making it widely used across various tasks.

Once the data is optimised, it is split into separate training, validation, and test sets to effectively evaluate the model's performance and generalisability. This separation is crucial for checking how well the model performs on unseen data and ensuring that it does not simply memorise the training data. Typically, the training set is the largest, providing substantial data for the model to learn and generalise. The validation set, which is smaller, is used to fine-tune model parameters and make adjustments during training, while the test set, with the same size, is reserved for the final evaluation to assess how well the model generalises to entirely new data.

During the training process, the primary goal is to optimise a loss function, which measures the discrepancy between predicted and actual values. Gradient descent and its variants are used to minimise this loss, with the learning rate playing a crucial role in ensuring effective convergence. The loss function also helps quantify different types of errors, which can reveal issues related to bias and variance. Key errors to consider during training include human-level error, training error, and validation (or development) error. Human-level error represents the best achievable performance for the task, serving as a benchmark for comparison. Whereas, the training and validation errors refer to the loss function's performance on the training and validation sets respectively.

A significant gap between the human-level error and training error typically indicates an issue with bias, suggesting that the model is not learning the underlying patterns effectively. Conversely, a large difference between training and validation error often points to a variance issue, meaning the model may be overfitting to the training data. To manage bias, strategies such as using more complex models, extending training duration, and employing advanced optimisation algorithms can be effective. On the other hand, variance can be addressed by acquiring more data, applying techniques like regularisation and dropout, augmenting the data, and performing hyperparameter searches.

2.5.2 Testing

To assess the final performance of a machine learning model on unseen data, the test set comes into play. This set is used to compute the test error by applying the same loss function used during training. A significant discrepancy between the validation and test errors may indicate overfitting to the validation set. In such cases, increasing the amount of validation data or revisiting model tuning can help address this issue.

Evaluating a model's performance involves several key metrics and visualisations. One of the most informative tools is the Receiver Operating Characteristic (ROC) curve, which plots the true positive rate (TPR) against the false positive rate (FPR). The TPR, also known as recall, measures the proportion of actual positives correctly identified by the model. The FPR, on the other hand, indicates the proportion of false positives among all actual negatives. The ROC curve helps in understanding the trade-offs between sensitivity and specificity across different thresholds.

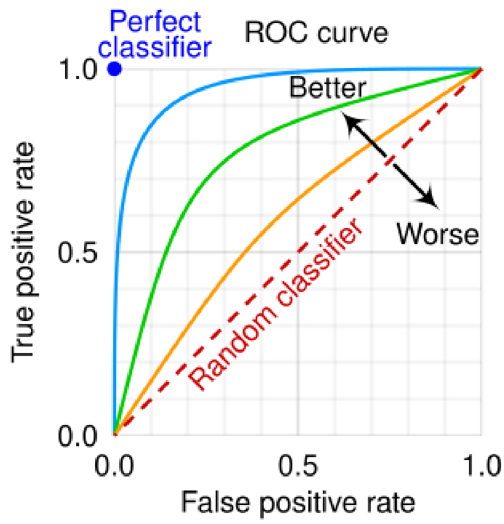


Figure 13: ROC curves for three models showing the trade-offs between the true positive rate (sensitivity) and the false positive rate across various thresholds [23]. The diagonal dotted line represents the baseline performance of a random classifier. The dot in the upper left corner indicates the ideal performance, with each curve demonstrating the model's performance improvements relative to each other.

Another important metric is precision, which quantifies the proportion of positive predictions that are truly positive. Precision differs from the false positive rate in that the false positive rate is calculated as the ratio of false positives to the total number of actual negatives. The F1 score, which is the harmonic mean of precision and recall, provides a single metric that balances the two, offering a comprehensive measure of model performance. It is calculated as

$$F_1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{2.6}$$

In addition to these metrics, plotting predicted values against actual values can offer valuable insights. This visualisation helps in identifying areas where the model may be underperforming or biased, revealing potential issues such as systematic errors or areas where the model struggles.

The choice of metrics ultimately depends on the specific goals of the model. For instance, in a medical diagnostic context, high recall may be prioritised to ensure that most positive cases are detected, even at the expense of precision. In contrast, in a spam detection application, precision becomes more critical to prevent legitimate emails from being misclassified as spam.

By leveraging these metrics and visualisations, one can comprehensively evaluate a model's performance, ensuring it meets the required standards for its intended application.

3 Methods

This thesis is divided into two main sections. The first focuses on the application of machine learning algorithms for the classification of signal versus background events in the fully hadronic VVjj channel (Fig. 3). The insights gained from mastering these algorithms and their implementation lay the groundwork for the second section, which centres on the development of the polarisation tagger. The progression from the initial algorithmic analysis to the practical implementation highlights the continuous advancement of expertise and application throughout the thesis.

3.1 Signal vs. Background Classification

This section establishes the foundational methods for distinguishing between signal and background events, focusing on feature analysis and machine learning models. The results and methodologies developed here serve as a basis for the subsequent section on polarisation tagging, where the principles of classification are adapted to tackle the more nuanced challenge of identifying different polarisation states.

3.1.1 Feature analysis

The first step in signal vs. background classification is to analyse the data to understand its nature and distribution. A thorough feature study offers valuable insights into the types, ranges, and statistical properties of the features, which are crucial for designing an effective classification algorithm. For this study, simulated Monte Carlo (MC) data is employed, specifically the mc16a_16feb23 and mc16e_16feb23 datasets, covering both signal and background events. For background events, simulated data from QCD dijet, V+jets and VV processes are used.

The signal events under consideration span various mass ranges of the Radion Model, Heavy Vector Triplet (HVT) Model [18], and Bulk Randall-Sundrum Model [12]. Initially, the distributions of one selected signal, the Radion model with a mass of 2000 GeV, are compared to those of the backgrounds. The specific signal mass is not critical to the analysis, but 2000 GeV is chosen because it coincides with the peak of the background distribution (Fig. 14). Optimising the discrimination between signal and background at this mass represents a baseline achievement in distinguishing the two.

Various jet features are explored to identify those with the highest discriminative power between signal and background, focusing on the primary-, secondary-, and dijet information. By pinpointing the most predictive features and identifying those that are redundant or irrelevant, a more informed feature selection process can be undertaken. Training the model using only these important features preserves performance while reducing computational complexity.

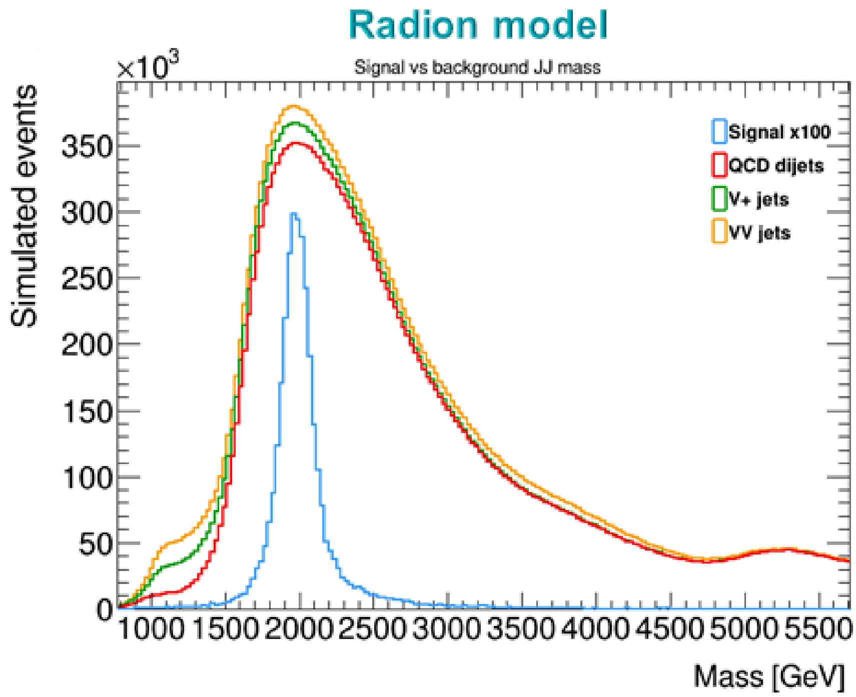


Figure 14: Comparison of dijet (JJ) mass distributions for the 2000 GeV Radion signal and stacked background processes. The signal events are enhanced by a factor of 100 to highlight their difference relative to the total background contribution. The signal peak around 2000 GeV overlaps with the background peak, which is predominantly composed of QCD dijet events. The additional peaks observed in the QCD dijet distribution arise from the background simulation conducted across various dijet mass slices, resulting in a non-continuous distribution with noticeable fluctuations. Nevertheless, this deviation has minimal impact on the classification task.

After analysing a specific signal signature, the distributions of multiple signals are compared to evaluate their consistency with expected patterns. Additionally, the distributions of the models across various mass ranges are examined. This comparison assesses whether the classification model performs consistently across different signals and how its effectiveness varies with each signature.

3.1.2 Binary classification

Once the feature analysis is completed, the next step is model design and training. To estimate the human-level error in the signal vs. background classification problem, a boosted decision tree model is selected for its balance of simplicity and strong classification performance. In particular, XGBoost is chosen because of its straightforward implementation, which facilitates efficient model training. Additionally, a binary classification algorithm employing logistic regression is employed to predict the probability of a data point being a signal. The performance of this model is evaluated using the Log-loss function, a standard metric for binary classification that quantifies the accuracy of probability predictions.

The model is optimised through systematic fine-tuning of various hyperparameters. Starting with a relatively high learning rate to accelerate training, it is gradually reduced to avoid overfitting. Similarly, the tree depth is initially set to five and incrementally increased to balance performance with model complexity. Subsampling and column sampling rates are fixed at 0.8 to further mitigate overfitting. On top of that, early stopping is also applied, beginning with three rounds without improvement and adjusted upwards to assess its impact on performance. To ensure reproducibility, a random seed of 42 is used. After training, the model's effectiveness is evaluated using several metrics, including the ROC curve, Precision-Recall (PR) curve, F1 score, Approximate Median Significance (AMS), and accuracy.

Once a performance benchmark is established with the BDT, attention shifts to achieving comparable results with a deep neural network. The DNN is selected for its ability to capture complex patterns and abstractions, making it particularly suitable for the polarisation tagger. Specifically, PyTorch is selected for its flexibility and ease of modification, enabling efficient experimentation and tuning of the model.

The exploration of different neural network architectures begins with a shallow model consisting of a single hidden layer with 10 neurons. After training and evaluating this model, larger and more complex architectures are tested by gradually adding neurons and hidden layers, experimenting with different configurations. Each model is trained and tested independently, and adjustments are made between sessions based on performance. Once increasing the model's complexity no longer results in improved performance, the least complex model that performs best is selected. Throughout this process, the ReLU activation function is applied for its ability to introduce non-linearity while mitigating the vanishing gradient problem, which is advantageous for deeper networks [24]. In the output layer, the Sigmoid function is utilised to generate probabilities for binary classification. To ensure consistency with the BDT approach during optimisation, the Log-loss function is employed as well.

Similar to the BDT, the learning rate is fine-tuned to ensure efficient convergence without overshooting. Additionally, the Adam optimiser is used for its adaptive learning rate properties, further enhancing training efficiency. Regularisation is applied with varying strengths to address different concerns: small values (10^{-6} to 10^{-4}) are used when the model risks underfitting, moderate values (10^{-4} to 10^{-2}) help control overfitting without significantly affecting performance, and larger values (10^{-2} to 10^{-1}) are implemented when overfitting becomes a major concern. If the model does not improve during training, the learning rate is halved, and the regularisation strength is doubled to mitigate overfitting. Early stopping is then applied to halt training after three epochs without improvement in validation loss. To ensure a comprehensive assessment after training, the DNN's performance is rigorously evaluated using the same metrics as the BDT.

3.2 Polarisation tagging

Building upon the foundational methods established above, the following section delves into the development of a polarisation tagger. While classification principles provide a solid base, polarisation tagging introduces additional complexities in distinguishing between longitudinally and transversely polarised vector bosons. To address these challenges, a controlled environment is employed, initially focusing on evaluating the feasibility of polarisation distinction and refining the tagging approach. The section explores the unique features and machine learning models tailored to this specialised task, highlighting how techniques are adapted to meet the specific requirements of polarisation tagging. More complex scenarios, such as background interference, will need to be investigated in future research.

3.2.1 Feature analysis

In line with the approach taken in signal versus background classification, the initial step in developing the polarisation tagger is to conduct a thorough feature analysis. This process involves examining the data to understand its characteristics and distribution. By doing so, valuable insights are gained into the features that will guide the design of the polarisation tagger.

For this study, pure Monte Carlo (H5 files from June 2024 based on MC20[1 2]) data for the process W'→WZ→ qq̄qq̄ is used. This dataset includes jet, subjet, and particle flow (pflow) track information for both transversely and longitudinally polarised vector bosons, generated using leading-order (LO) MadGraph. By focusing on this specific dataset, a detailed exploration of the features relevant to distinguishing between different polarisation states is made possible.

The analysis involves a comprehensive evaluation of each feature across jets, subjets, and the highest p_T particle flows. This thorough examination aims to identify which features provide the highest discrimination between longitudinally and transversely polarised vector bosons. In addition, feature engineering is employed to enhance the distinguishing power of the tagging approach, guided by theoretical considerations and expected outcomes. Given that polarisation is closely related to the decay angle between the vector boson and its decay products in the rest frame, features that approximate this angle or its cosine are anticipated to improve discrimination between the polarisation states.

This work considers four estimations of the cosine of the decay angle. The first estimation utilises Equation 2.4 to approximate $\cos\theta$, allowing for direct inference of the angle in the rest frame of the vector bosons from lab-frame observables. The remaining three estimations are

[1] mc20_13TeV.426349.Pythia8EvtGen_A14NNPDF23LO_WprimeWZ_flatpT_longpol.deriv.DAOD_JETM2.e6961_s3797_r13144_p5707

[2] mc20_13TeV.426350.Pythia8EvtGen_A14NNPDF23LO_WprimeWZ_flatpT_transpol.deriv.DAOD_JETM2.e6961_s3797_r13144_p5707

based on the dot product of two vectors, expressed as $\vec{a}\cdot\vec{b} = |\vec{a}||\vec{b}|\cos\theta$. These include simplified calculations in the transverse plane of the lab frame, specifically $\cos\theta = \cos\Delta\phi$, as well as three-dimensional calculations that incorporate the p_z components. Additionally, a rest-frame approximation of the angle between the two subjets is derived using Lorentz transformations, with the boost factor obtained from the momentum of the vector boson. A detailed derivation of these approximations can be found in Appendix A: Cosθ dot product estimations.

3.2.2 Binary classification

Successful completion of the feature analysis is followed by the design and training of machine learning models. A BDT, implemented with XGBoost, is used to predict binary probabilities between longitudinally and transversely polarised states, with logistic regression facilitating the classification. The model's performance is evaluated using the Log-loss function, and hyperparameters are systematically fine-tuned, following the same approach used in the signal versus background classification task.

After establishing a benchmark with the BDT, attention shifts to training a deep neural network, chosen for its ability to capture intricate patterns and abstractions vital to polarisation tagging. Given the complexity of the task, the initial DNN architecture consists of three hidden layers, each containing 100 neurons. The model is then expanded iteratively by adding neurons and layers, exploring various configurations to enhance performance. Once further increases in complexity no longer yield improvements, the most efficient architecture is selected. ReLU activation functions are applied in the hidden layers, while a Sigmoid function generates binary probability outputs. For consistency in validation, the model's effectiveness is assessed using the same set of metrics applied to the earlier approaches.

3.2.3 Cosθ regression

Since polarisation classification is inherently challenging, especially in the fully hadronic channel, additional methods beyond binary classification are valued. One promising approach involves using a regression model to accurately approximate parton-level values, thereby enhancing the discrimination between polarisation states. The most intuitive target for regression is $\cos\theta$, as it directly correlates with polarisation (Fig. 6). Various estimations of this angle have already been explored in the feature analysis above. For a more accurate and discriminative use of $\cos\theta$, however, a refined approach is necessary. A machine learning regression model provides a viable path to achieve this. To eliminate generator dependence, regression can be performed on $\cos\theta$ reconstructed using exclusive k_T subjets (with $n_{\text{subjet}} = 2$) in the rest frame of the Large-R jet, as this aligns well with the parton-level $|\cos\theta|$ distribution (Fig. 15).

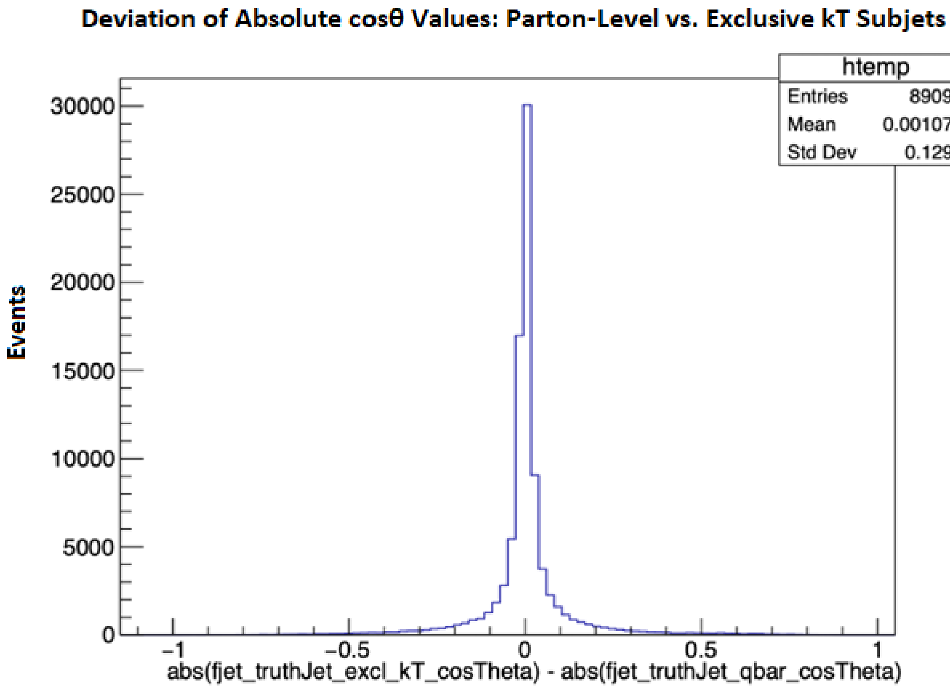


Figure 15: This plot illustrates the deviation between the absolute values of $\cos\theta$ at the parton level and those reconstructed using exclusive k_T subjets [25]. The comparison highlights the effectiveness of exclusive k_T subjets in approximating parton-level $\cos\theta$ values, which is crucial for accurate particle identification and analysis in the context of polarisation tagging. The distribution centres nicely around zero and exhibits a standard deviation of 0.13, indicating the resolution and spread of the reconstructed $\cos\theta$ values.

The regression model employed in this thesis is built upon the Weaver framework, a streamlined yet flexible machine learning tool specifically designed for high-energy physics (HEP) applications [26]. Prof. Dr. Shudong Wang has employed this framework to develop the constituent-based W tagger [27], presenting a promising step toward achieving constituent-based polarisation tagging.

To prepare the input data for the model, normalisation is performed using the StandardScaler from the sklearn.preprocessing module. This process involves fitting the scaler to the training data and transforming it, ensuring that all features are on a comparable scale. By normalising the data, the model can focus on learning the underlying patterns without the added complexity of adjusting for different feature scales. Subsequently, the same transformation is applied to the validation and test datasets, maintaining consistency across all datasets and enhancing the model's learning efficacy. The training samples[3] used in this process are derived from the ConstituentTaggers datasets[4], which ensures a robust foundation for training.

With the data appropriately normalised, the next step involves the ParticleTransformer model. This model integrates both dense and transformer-based layers to perform regression tasks on particle physics data. It begins by embedding the input features into a higher-dimensional space using three fully connected layers. Each of these layers incorporates GELU activation functions. This embedding serves as the initial transformation of the input before it is processed by the transformer's attention mechanism. In deep and complex architectures, GELU activation functions can enhance performance due to their smoother and probabilistic characteristics. These features are advantageous for capturing intricate patterns within the data.

The core of the model is an 8-layer transformer with eight attention heads per layer. This transformer architecture allows the model to capture intricate relationships between different particle features by employing self-attention. The use of multiple encoder and decoder layers further enhances the model's ability to model both global and local dependencies within the input sequence.

The output of the transformer is passed through a final set of fully connected layers that progressively reduce the dimensionality, again utilising GELU activation to introduce non-linearity. The model's final output is generated using only the first token (i.e., the first element of the sequence), which is passed through a regression head to produce predictions. This output layer is specifically designed for regression tasks, enabling the model to generate continuous values that represent the predicted variable, such as $\cos\theta$.

During training, the Mean Squared Error (MSE) loss function is utilised to quantify the difference between predicted and true values. MSE is particularly well-suited for regression tasks, as it penalises larger deviations more heavily, encouraging the model to prioritise minimising significant errors.

The evaluation of the regression model follows a similar approach to that of the previous deep neural network model. However, since the metrics applicable to classification do not apply here, the assessment centres on the deviation between predicted and true values. This methodology ensures a robust evaluation of the model's performance in accurately predicting continuous variables.

[3] /eos/atlas/unpledged/group-tokyo/users/tnobe/works/JSS/training/constituent_based_w_tagging/preprocessing/root/output_WJet_or_ZJet.root.

[4] /eos/atlas/atlascerngroupdisk/perf-jets/TAGGING/TopBosonTagAnalysisRel21/FlattenerOutput/ConstituentTaggers.

4 Results

The results section mirrors the structure of the methods section, comprising two main parts. The first part presents findings from the study focused on classifying signal versus background events in the fully hadronic VVjj channel (Fig. 3). It begins with the results of the feature analysis, highlighting the most significant features for classification tasks and illustrating the differences between various event signatures. Following this, the application of machine learning algorithms is discussed, demonstrating their effectiveness in distinguishing between signal and background events.

The second part centres on the outcomes of the polarisation tagger, beginning once more with results from the feature analysis. This is followed by an evaluation of the model's ability to accurately identify the polarisation states of vector bosons. Finally, the results of the $\cos\theta$ regression model are presented, detailing its performance in estimating truth-level values. This structured approach facilitates a clear presentation of the results obtained from both classification and polarisation tagging processes.

4.1 Signal vs. Background Classification

4.1.1 Feature analysis

Following the feature analysis of signal versus background events in the fully hadronic VVjj channel, the most discriminative features are identified, offering valuable insights into the characteristics, ranges, and statistical properties essential for an effective classifier.

Initially, the distributions of the selected Radion model signal (m = 2000 GeV) are compared against the backgrounds. Various jet features, including primary-, secondary-, and dijet-related information, are examined to assess their discriminative power. The most predictive features are illustrated in Figures $16_{\mathrm{a,b,c,d}}$, which include the DisCo score, jet mass, number of tracks, and the D2 score of the primary jet. The DisCo score serves as a regularisation term based on distance correlation, quantifying non-linear correlations between the Energy Flow Polynomials (EFPs) and the reference labels [28]. This makes it particularly effective for identifying complex correlations between multiple features and the signal, achieving state-of-the-art performance in tasks such as W tagging. In contrast, the D2 score represents a two-prong discriminant based on a jet shape variable [29], distinguishing between one- and two-prong jets by producing small values for two-prong jets and larger values for one-prong jets.

The analysis reveals that secondary jet features offer minimal additional discrimination. Similarly, most dijet features do not enhance the model's ability to distinguish between signal and background, with the exceptions being energy and mass. These characteristics relate closely to the heavy resonance decaying into two vector bosons, as illustrated in Figures $16_{\mathrm{e,f}}$.

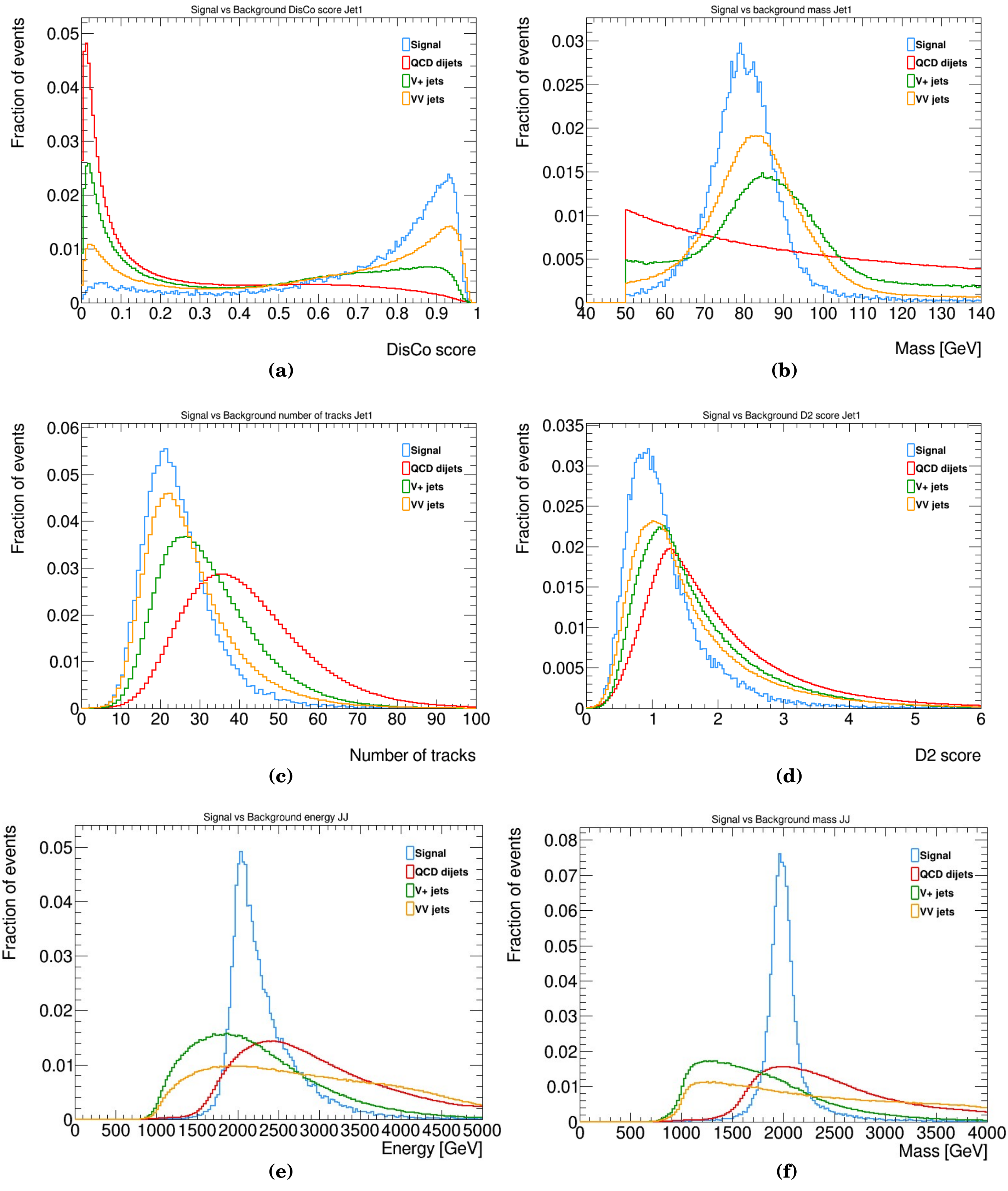


Figure 16: Distributions of the most discriminative features for the Radion model signal (m = 2000 GeV) compared to various background processes. Panels (a) to (d) illustrate primary jet features, including the DisCo score, mass, number of tracks, and D2 score, all of which aid in distinguishing the signal from background. Panels (e) and (f) present the predictive dijet features, specifically energy and mass distributions.

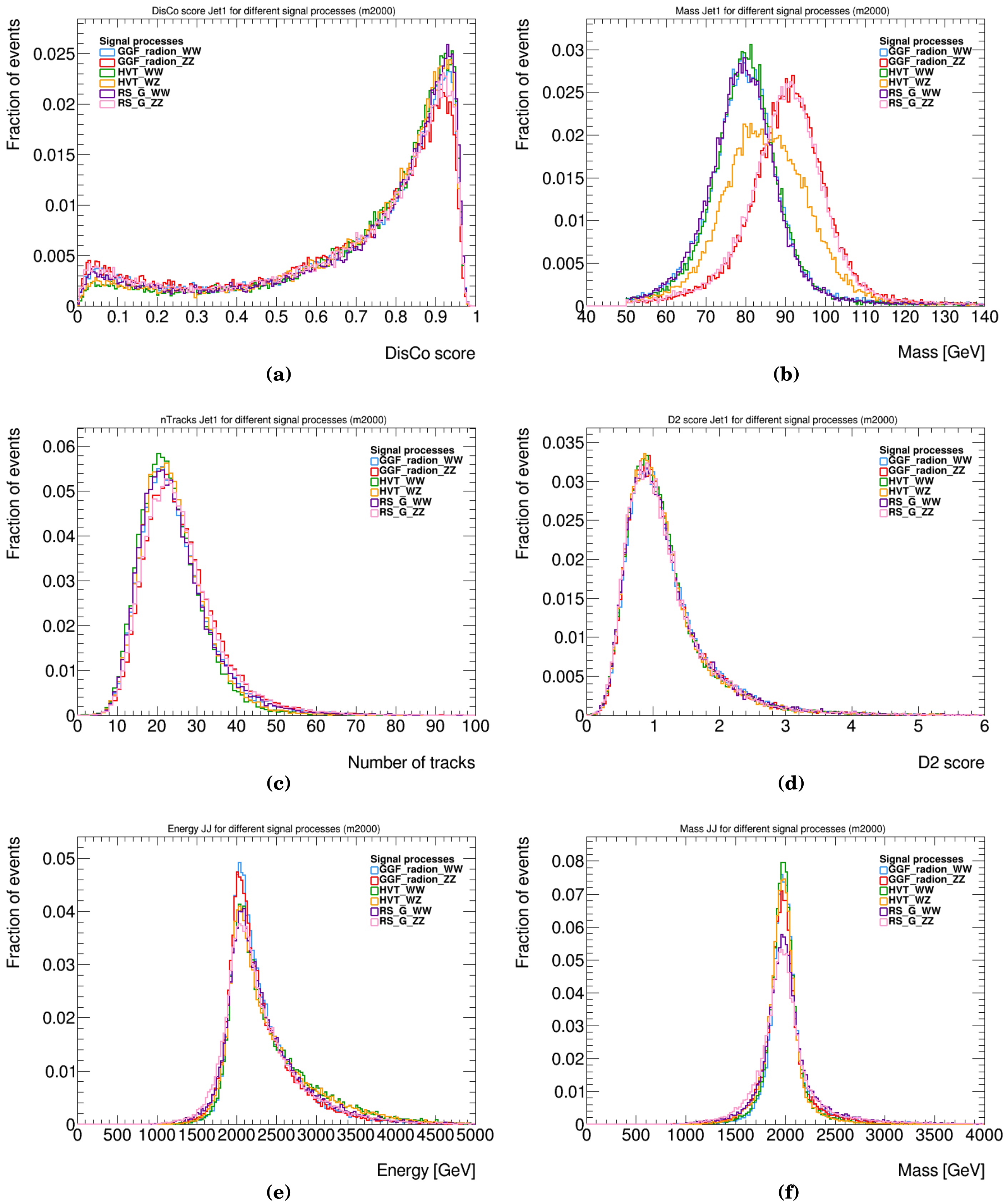


Figure 17: Distributions of the most discriminative features for various signal models, including the Radion Model, Heavy Vector Triplet Model, and Bulk Randall-Sundrum Model. Panels (a) to (d) present primary jet features, while panels (e) and (f) illustrate dijet features of energy and mass. These features remain relatively consistent across the models, with minor variations in jet mass reflecting the different vector boson configurations (WW, WZ, ZZ) associated with each signal.

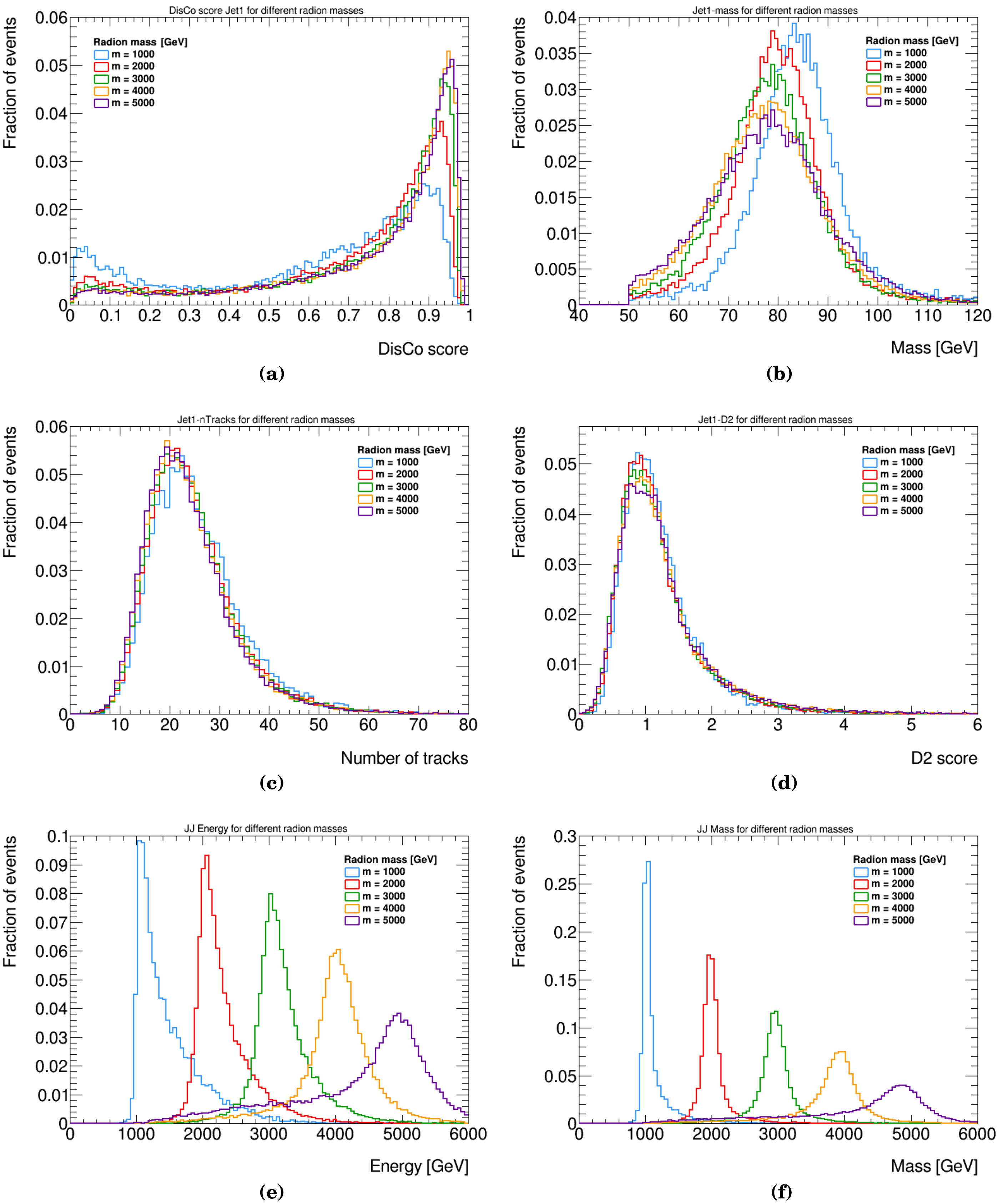


Figure 18: Distributions of the most discriminative features for the Radion model signal at various resonance masses. Panels (a) to (d) demonstrate that the primary jet features, including the DisCo score, mass, number of tracks, and the D2 score, remain consistent across different resonance masses. In contrast, panels (e) and (f) present the dijet features of energy and mass, which scale with the resonance mass.

After analysing a specific signal signature, the distributions of multiple signals are compared in Figure 17 to evaluate their consistency with theoretical expectations. The signals considered include the Radion Model, Heavy Vector Triplet Model, and Bulk Randall-Sundrum Model. The most discriminative features remain consistent across these various signal signatures, with minor differences noted in the jet mass. This variation reflects the differing vector boson configurations (WW, WZ, ZZ) associated with each model, peaking around their respective vector boson masses.

Furthermore, examining the signal distributions across different mass ranges allows for assessing the consistency in discrimination power across various signatures, offering insight into potential variations in classification performance. Figure 18 shows how the key features change across different resonance masses. While the most discriminative jet features remain consistent, the dijet features of energy and mass scale with the resonance mass. This relationship is further illustrated by the energy and transverse momentum of the vector bosons into which the heavy resonance decays, as depicted in the jet distributions in Figure 19. Although these features exhibit limited discriminative power for lower resonance masses (m = 2000 GeV), their significance increases at higher masses (m > 3000 GeV).

Additional feature distributions, not shown in this section, include primary, secondary, and dijet features related to signal versus background, along with comparisons across different signals and mass ranges. These can be found in Appendix B: Signal vs. Background Plots.

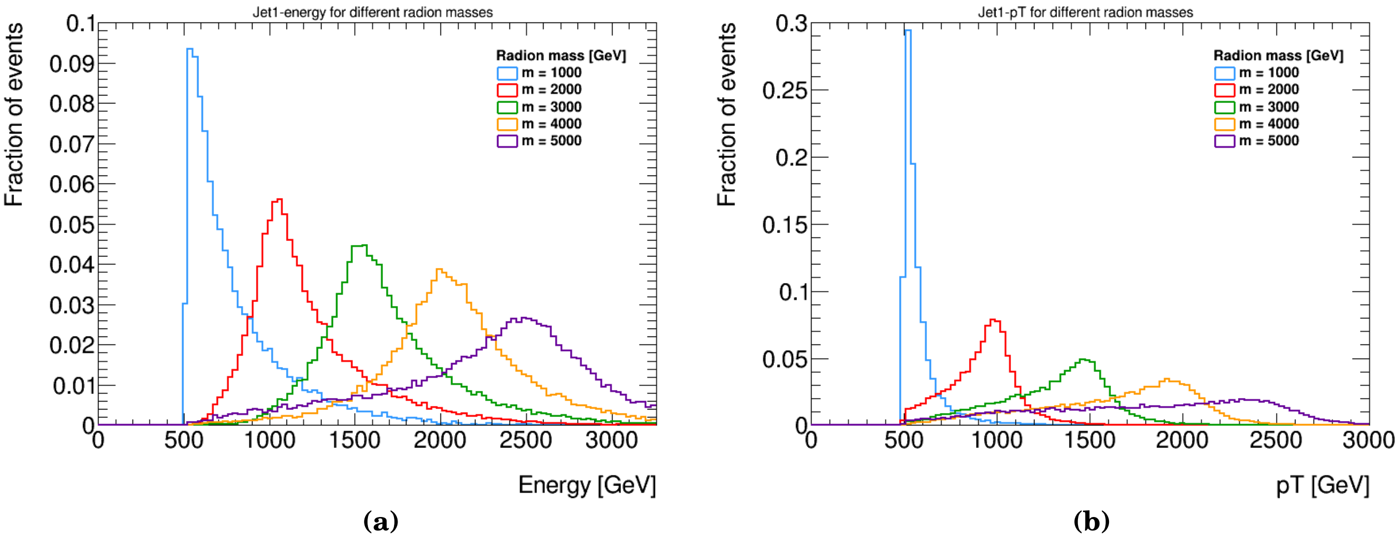


Figure 19: Distributions of primary jet features, energy, and transverse momentum (p_T) for the Radion model signal at various resonance masses. These features scale with resonance mass, highlighting their correlation with the vector bosons that decay directly from the heavy resonance.

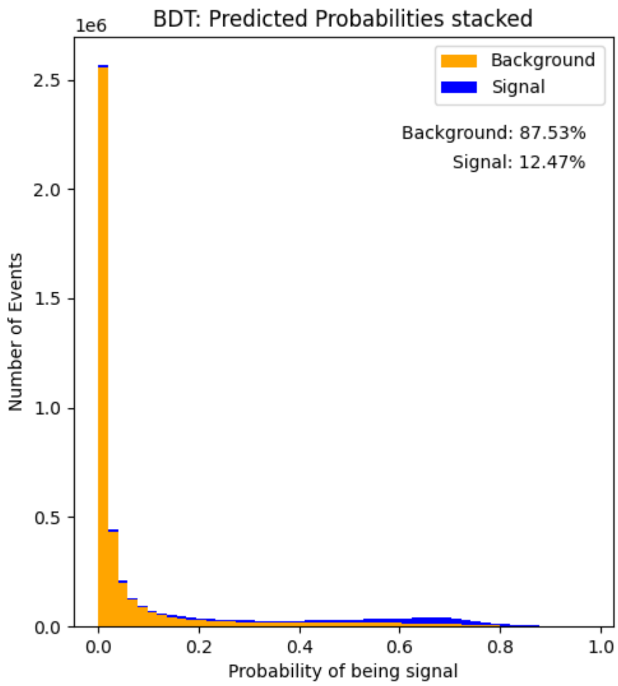
4.1.2 Binary classification

The results of the classification models, including boosted decision trees and deep neural networks, are presented in terms of predicted probabilities and performance metrics. A comparative analysis of their predictions and evaluation metrics highlights each model's effectiveness in distinguishing between signal and background events.

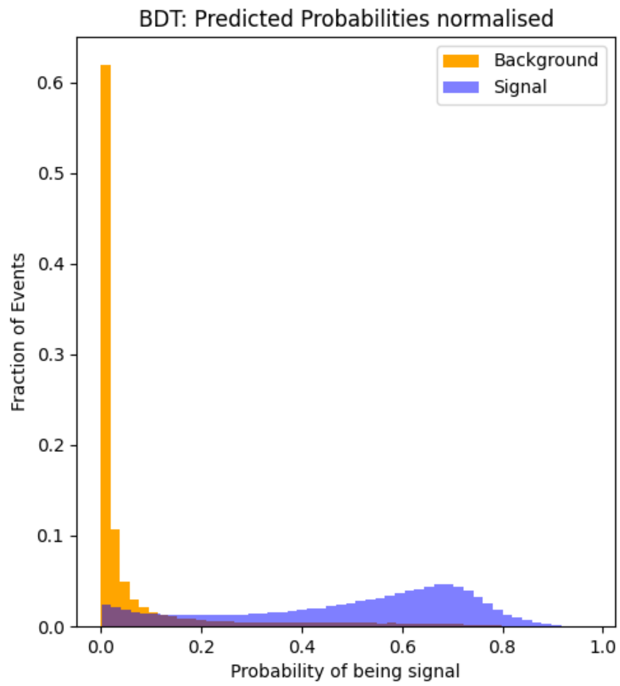
The best configurations of the models used in this thesis are as follows. The boosted decision tree model, with the objective of binary logistic regression, has a maximum depth of 10. This depth allows for a nuanced understanding of complex relationships within the data. The model parameters are set to a learning rate of 0.1, which governs the step size at each iteration while minimising the loss function. Additionally, subsampling and column sampling rates of 0.8 are employed to enhance model robustness by reducing overfitting. The BDT is trained for a maximum of 10,000 rounds and employs early stopping after 10 rounds without improvement, thereby ensuring that training is stopped when no significant progress is observed. The model's performance is evaluated using the Log-loss function, which quantifies the accuracy of predicted probabilities against the true labels.

In contrast, the best deep neural network model consists of two hidden layers, each containing 100 neurons and utilising the ReLU activation function. This architecture allows the model to learn intricate patterns in the data through non-linear transformations. The output layer features a single node with a sigmoid activation function, which is ideal for binary classification tasks as it outputs probability values between 0 and 1. The learning rate is similarly set at 0.1, complemented by a regularisation strength of 0.001, which helps mitigate overfitting by penalising excessively large weights in the network. The Adam optimiser is employed for training, known for its efficiency in handling sparse gradients and its adaptive learning rate capabilities. Like the BDT, the DNN uses log loss as the loss function. However, training is conducted for only 10 epochs, with early stopping activated after 3 epochs, allowing for quick convergence and efficient training cycles.

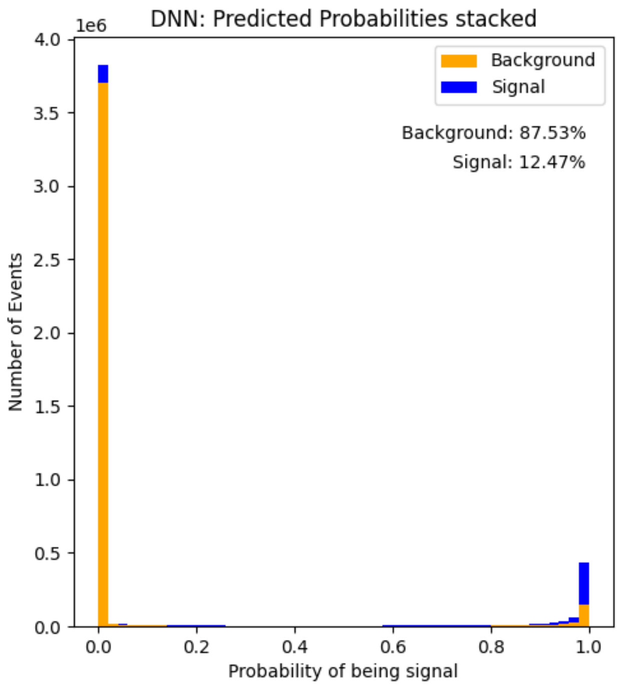
The predicted probabilities for both best-performing classifiers are illustrated in Figure 20. The stacked distributions in panels (a) and (c) clearly show a separation between background and signal events, with a greater number of signal events predicted at higher probabilities compared to background events. This separation indicates that both models excel at differentiating between the two classes, showcasing their potential for practical applications in event classification. The normalised distributions in panels (b) and (d) further emphasize the proportions of signal and background events, normalised to their respective counts. This highlights the fraction of predicted signal and background events across various probability thresholds, illustrating the models' effectiveness in accurately classifying events.



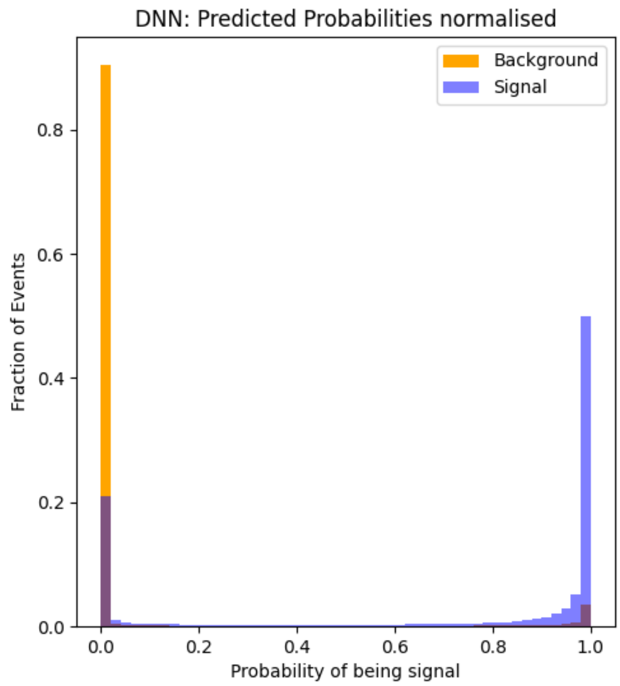
(a)



(b)



(c)



(d)

Figure 20: Comparison of predicted probabilities for the BDT and DNN classifiers. Panels (a) and (c) show stacked distributions of the predicted probabilities, highlighting the contributions from background (orange) and signal (blue) events. The background comprises 87.53% of events, while the signal represents 12.47%. Panels (b) and (d) present the normalised distributions, illustrating the fraction of events in relation to their respective total.

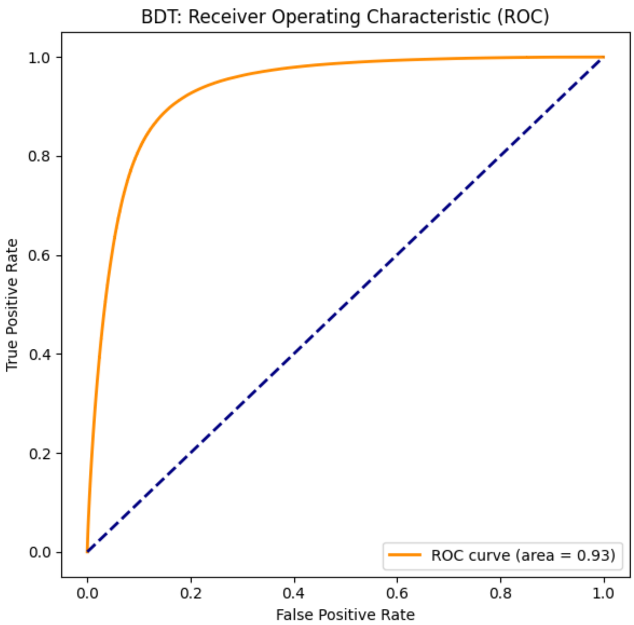
To further assess the performance of the BDT and DNN, various metrics are employed, including the ROC and Precision-Recall curves. The ROC curves, displayed in panels (a) and (c) of Figure 21, exhibit areas under the curves (AUC) of 0.93 for the BDT and 0.90 for the DNN. These values indicate strong discriminatory power, suggesting that both models are well-suited for the classification task at hand. In particular, the high AUC for the BDT suggests it has a slight edge in its ability to distinguish between signal and background events.

While the ROC curve is a useful tool for evaluating model performance, it can present an overly optimistic view in scenarios with a large number of background events, as is the case here. This occurs because the false positive rate can remain low, even with a considerable number of false positives, resulting in a high AUC that may not accurately reflect the model's practical performance. To address this limitation, the Precision-Recall curve is also employed. In contrast to the ROC curve, the PR curve focuses exclusively on the positive class, plotting precision against recall. This makes it particularly valuable for imbalanced datasets, where a large number of background events can obscure the model's ability to effectively identify signal events. Consequently, the PR curve offers a more transparent view of the model's performance in differentiating signal from background events.

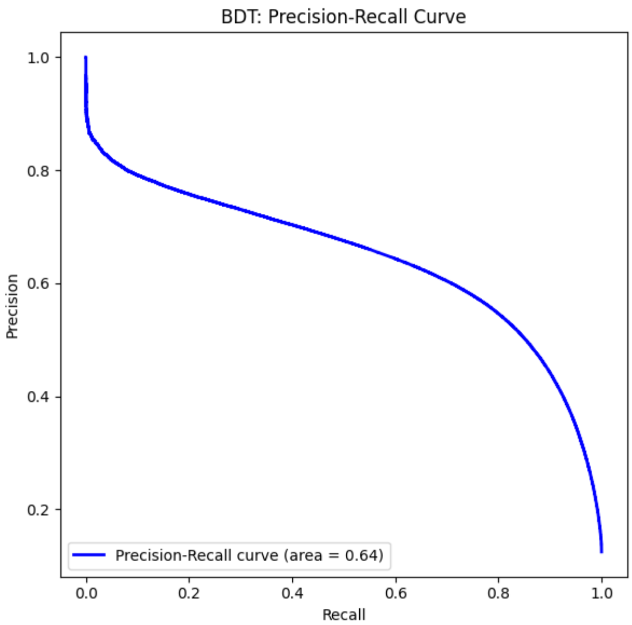
The Precision-Recall curves are shown in panels (b) and (d), illustrating the models' performance in terms of precision and recall across various probability thresholds. The areas under the PR curves are 0.64 for the BDT and 0.62 for the DNN. A perfect classifier would achieve a PR AUC of 1.0, indicating flawless precision and recall at all thresholds. On the other hand, a random classifier has a PR AUC equal to the proportion of positive labels in the dataset, in this case 0.12, indicating no better than chance performance. These metrics highlight the models' capability to accurately classify signal events while minimising false positives, which is essential for ensuring reliable results in particle physics analyses.

Supplementary performance plots for both the BDT and DNN models are available in Appendix B: Signal vs. Background Plots, providing additional validation and context for the results discussed here.

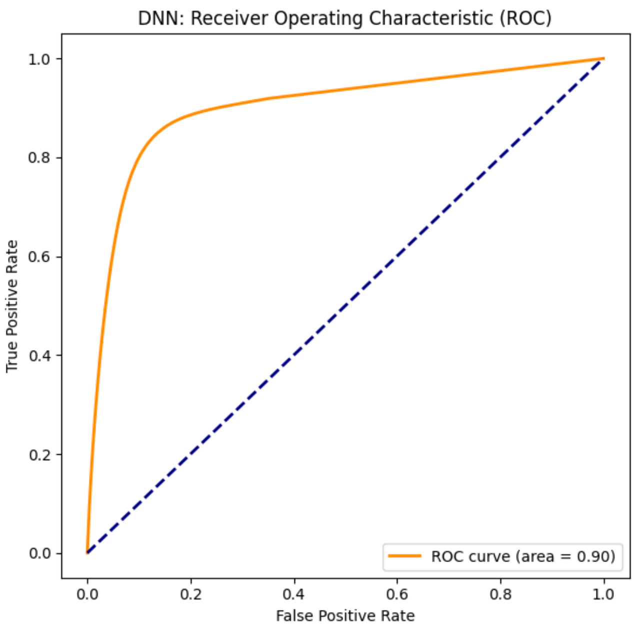
Overall, the results indicate that both the BDT and DNN models perform robustly in classifying signal versus background events. The BDT demonstrates slightly superior discriminatory power, while the DNN exhibits a more confident prediction pattern. These findings pave the way for further investigations into model refinement and feature engineering, ultimately contributing to improved classification techniques in the context of polarisation tagging.



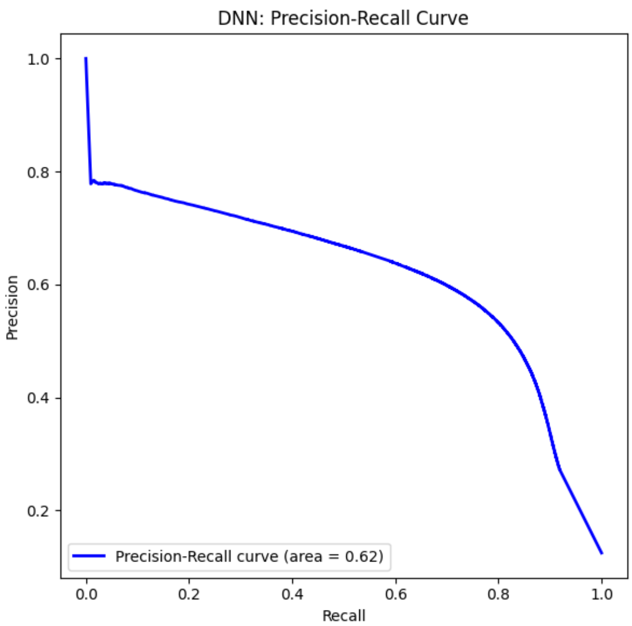
(a)



(b)



(c)



(d)

Figure 21: Comparison of performance metrics for the boosted decision tree and deep neural network models. The upper (a) and lower left (c) plots show the Receiver Operating Characteristic curves, with areas under the curves of 0.93 and 0.90, respectively, indicating strong discriminatory power. The upper right (b) and lower right (d) plots illustrate the Precision-Recall curves, with areas of 0.64 for BDT and 0.62 for DNN, highlighting the models' ability to distinguish between signal and background events across various probability thresholds.

4.2 Polarisation tagging

4.2.1 Feature analysis

In the classification of vector boson polarisation states, features from jets, subjets, and particle flow are examined to identify those that offer the most discrimination between longitudinally and transversely polarised bosons.

The most discriminative features are found in jet substructure variables, including $Z_{1,2}^{\mathrm{cut}}$, $\mathrm{Split}_{1,2}$, Angularity, and KtDR (Fig. 22). These variables provide valuable insights into the energy distribution and clustering behaviour of jets. $Z_{1,2}^{\mathrm{cut}}$ quantifies the energy sharing between the first two splittings, helping to distinguish how energy is allocated among the leading particles. $\mathrm{Split}_{1,2}$ measures the distance between jet constituents during the final clustering step, shedding light on the jet's internal structure. Angularity assesses the symmetry of energy distribution within the jet, revealing how energy deposits are arranged relative to the jet axis and aiding in identifying differences in jet shapes. Meanwhile, KtDR indicates the angular separation at the final clustering step, offering information about the spread or compactness of the jet in angular space. Together, these variables contribute to a comprehensive understanding of the underlying differences between the polarisation states of vector bosons, thereby enhancing the classification process.

After identifying the most discriminative jet substructure (JSS) features, the focus shifts to the subjet features themselves, which could further enhance the classification of vector boson polarisation states. As shown in Figure 23, several additional discriminative variables emerge, including ΔR, energy, and transverse momentum. The secondary subjet is particularly informative, typically showing lower energy and p_T values for transversely polarised vector bosons due to their decay kinematics. This is evident in the plot, where a greater number of secondary subjets are classified from longitudinally polarised vector bosons, showing higher energy and p_T values compared to their transversely polarised counterparts. Although ΔR alone offers limited variation, its combination with other features significantly enhances the performance of the classification models.

Additionally, the particle flow features are examined, yielding some discriminatory power. Given the significant computational demands of using the full pflow dataset, which can include up to 100 particle flows per event, the analysis focuses on only the four most energetic constituents. The most discriminative pflow features are illustrated in Figure 24, encompassing variables such as ΔR, $\Delta\phi$ and $\Delta\eta$. Furthermore, some discrimination is observed in flow energy and transverse momentum between the polarisation states, as shown in Figure 25.

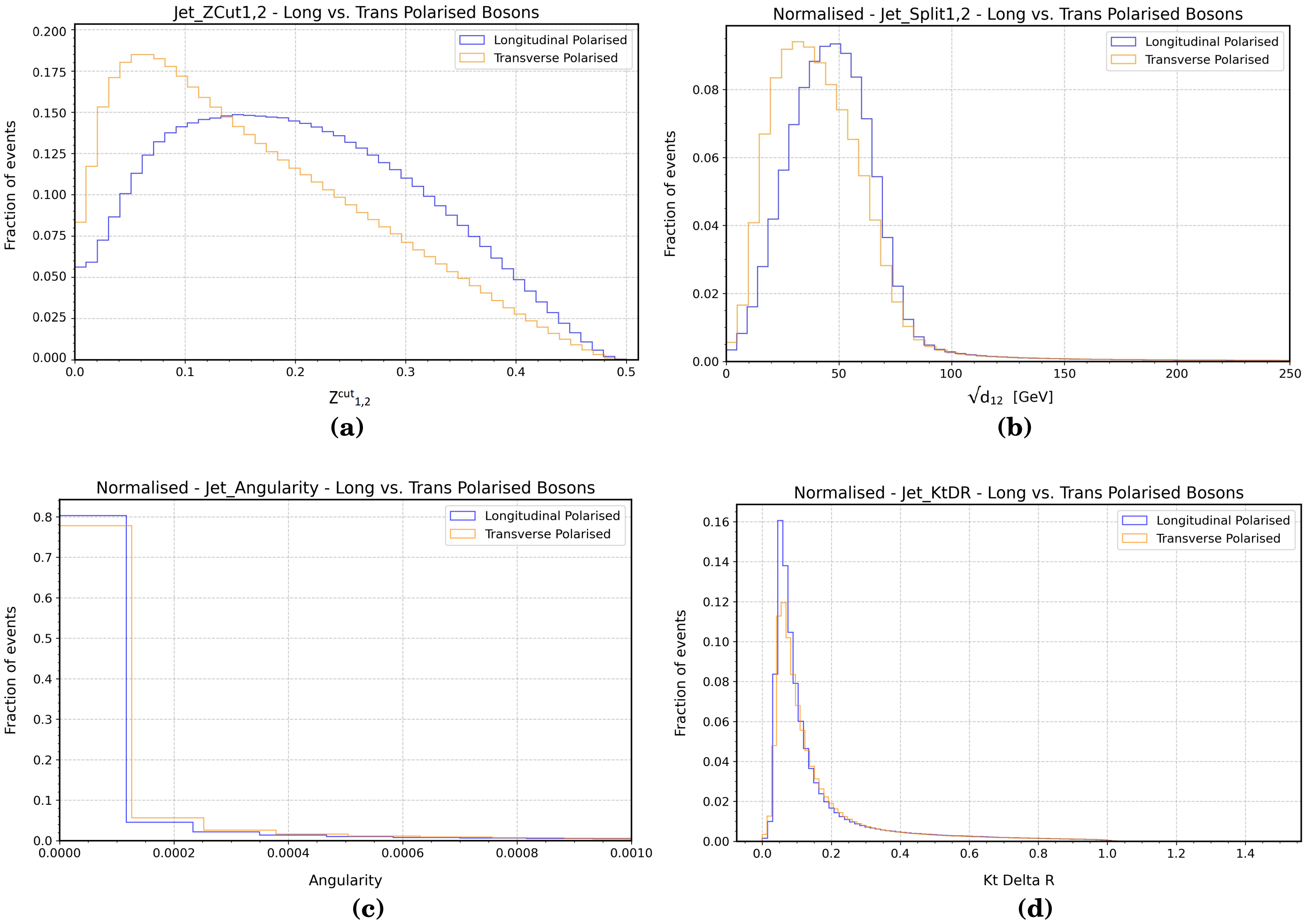


Figure 22: Distributions of the most polarisation-discriminative jet features, including the $Z^{\text{cut}}_{1,2}$, $\text{Split}_{1,2}$, Angularity, and KtDR. Notably, the first two features exhibit distinct distributions for longitudinal (blue) and transverse (orange) polarisation states. The $Z^{\text{cut}}_{1,2}$ parameter quantifies the energy sharing between the first two splittings, aiding in the distinction of energy allocation among leading particles. $\text{Split}_{1,2}$ measures the distance between jet constituents during the final clustering step, providing insight into the jet's internal structure. Angularity evaluates the symmetry of energy distribution within the jet, revealing how energy deposits are arranged relative to the jet axis and helping to identify differences in jet shapes. Lastly, KtDR indicates the angular separation at the final clustering stage, offering information about the spread and compactness of the jet in angular space.

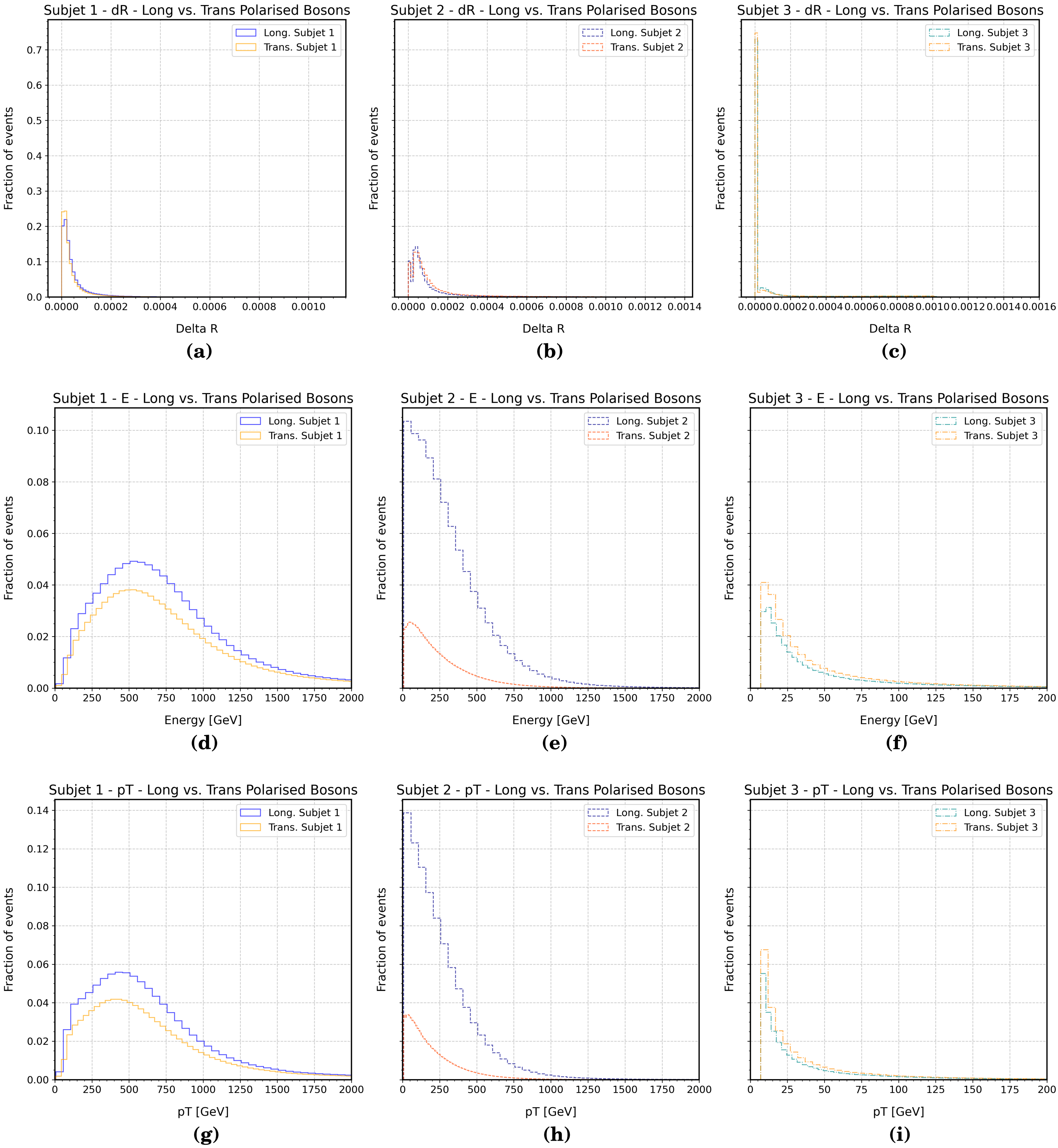


Figure 23: Distributions of the most discriminative subjet features for polarisation tagging, including ΔR, energy, and transverse momentum for all subjets, with the secondary subjet providing the highest discrimination between polarisation states.

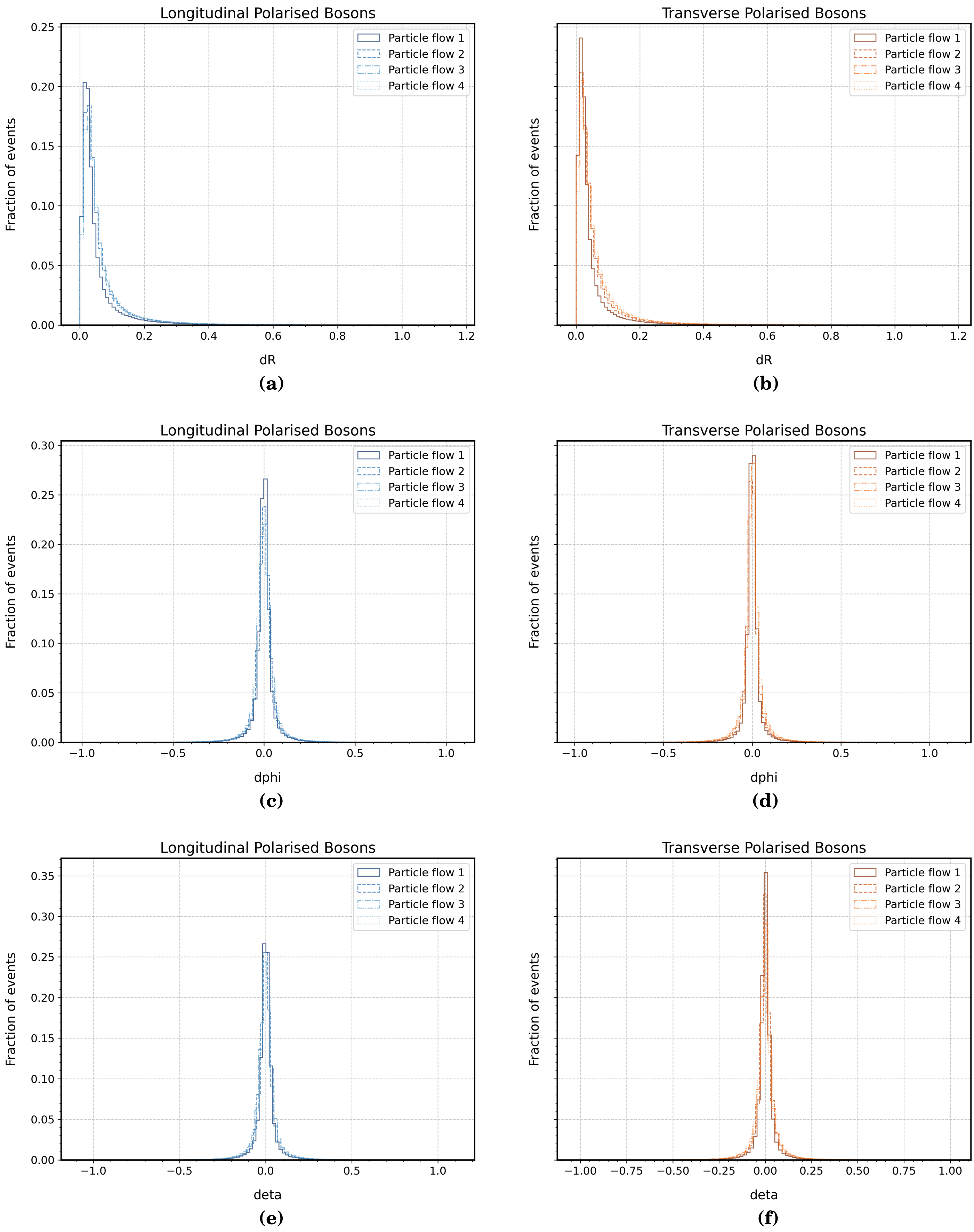


Figure 24: Distributions of the most discriminative particle flow features for polarisation tagging, including the flow ΔR, $\Delta\phi$ and $\Delta\eta$. The left plots depict the distributions for longitudinal polarisation, while the right plots show those for transverse polarisation.

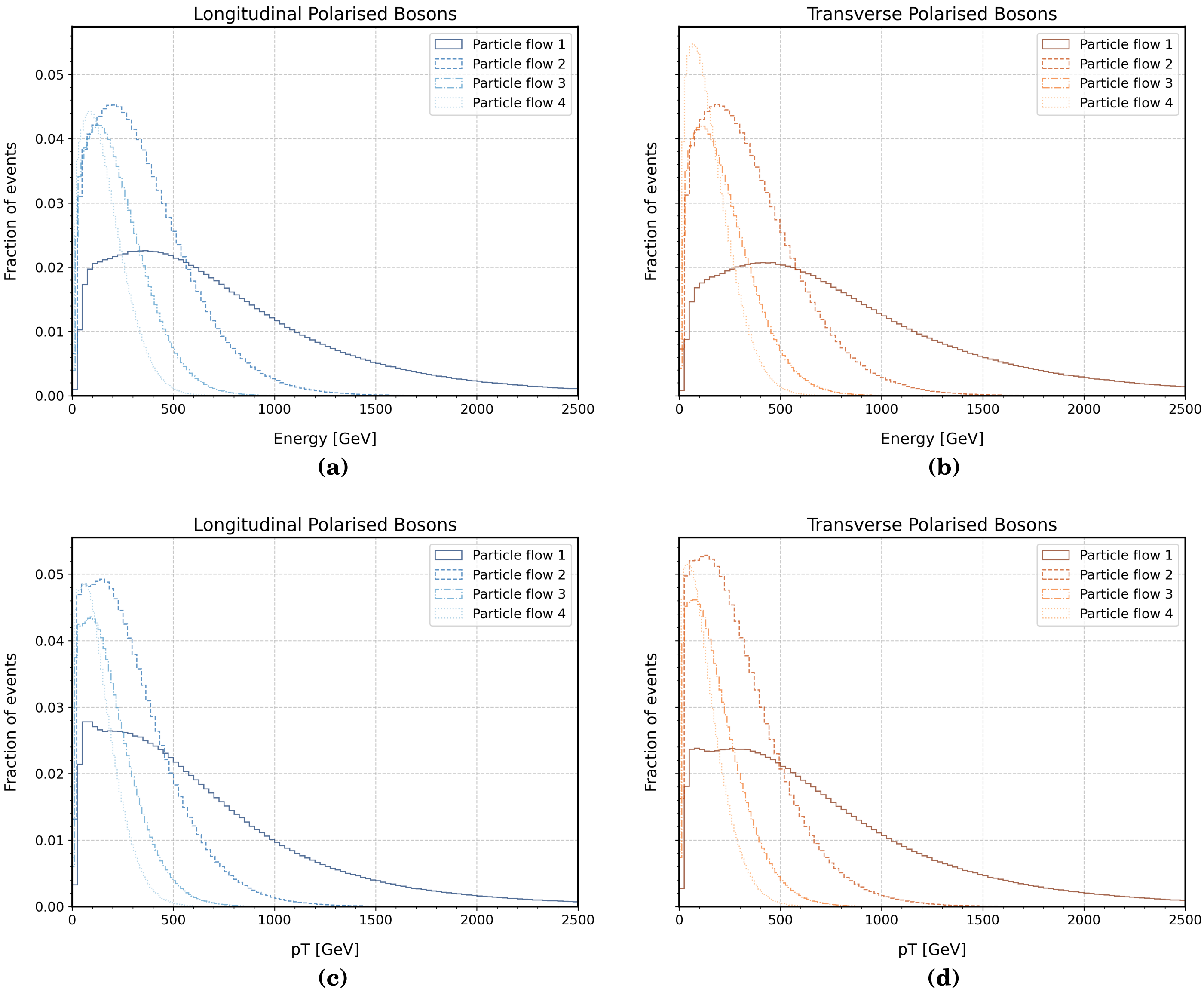
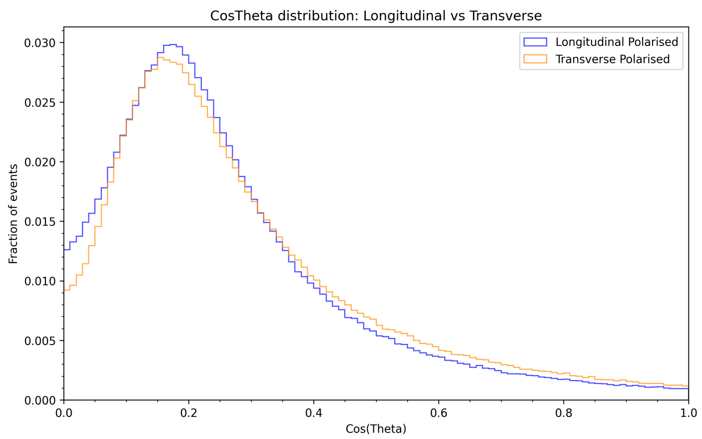


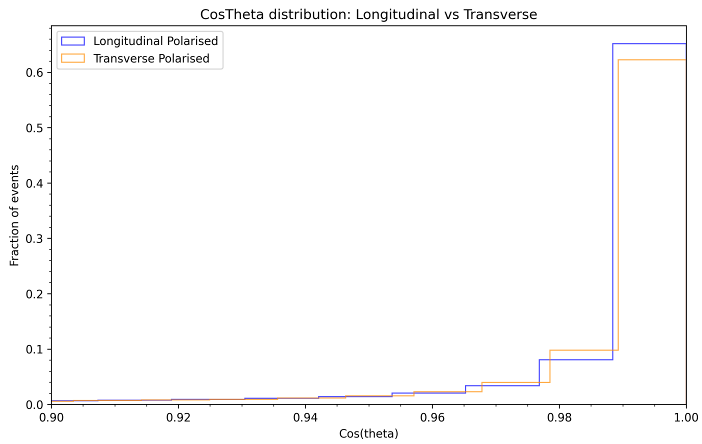
Figure 25: Distributions of additional discriminative particle flow features for polarisation tagging, including flow energy and transverse momentum. The left plots show the distributions for longitudinal polarisation, while the right plots depict those for transverse polarisation.

The final stage of the feature analysis includes additional engineered $\cos\theta$ features, as described in the methods section. These features are designed to enhance the discriminatory power of the classifiers. However, as shown in the plots (Fig. 26), they exhibit minimal differences between the polarisation states, indicating that further efforts are required to accurately estimate the decay angle.

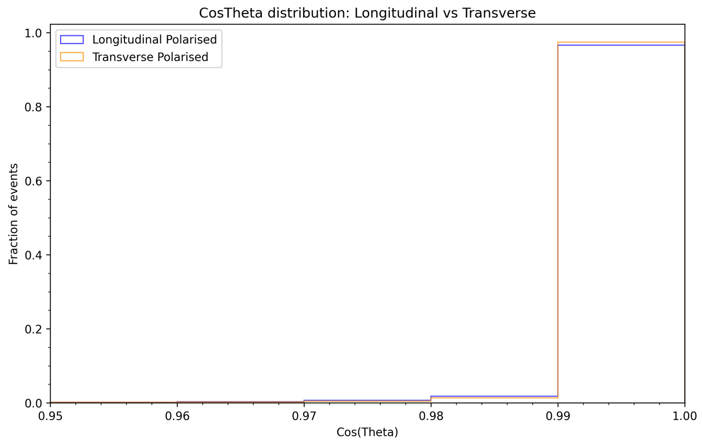
Additional feature distributions, not shown in this section, include jet, subjet, and particle flow features related to polarisation tagging, along with extra engineered features. These can be found in Appendix C: Polarisation Related Plots.



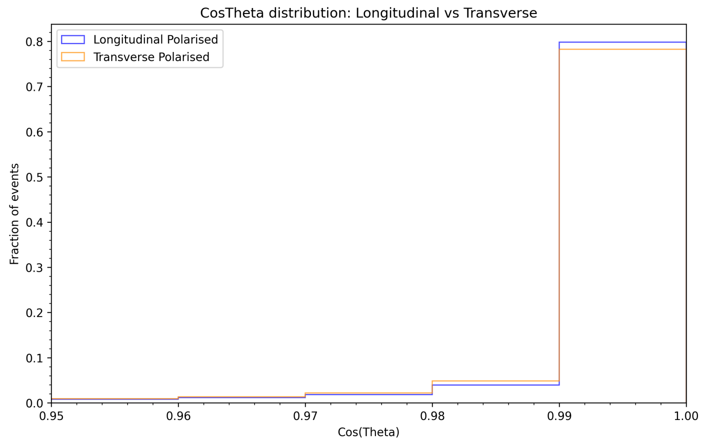
(a)



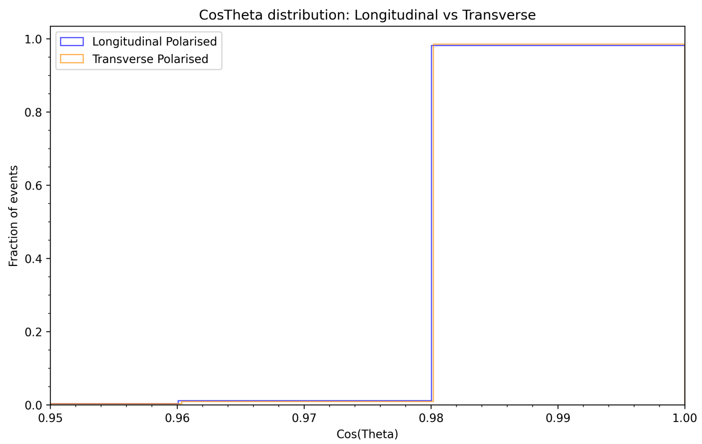
(b)



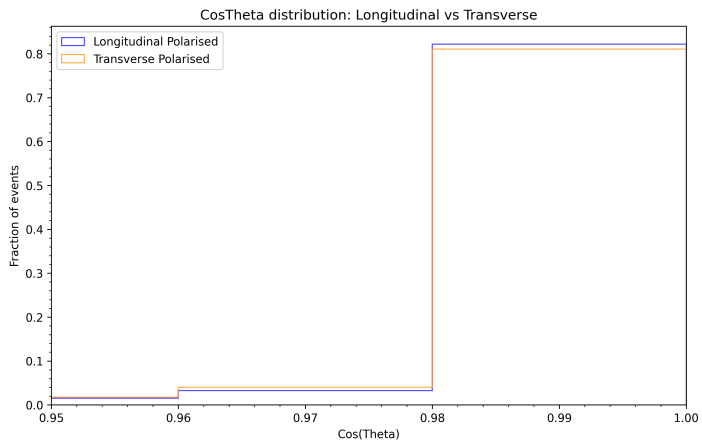
(c)



(d)



(e)



(f)

Figure 26: Distributions of the engineered $\cos\theta$ features for polarisation tagging. Panel (a) presents the results from applying Equation 2.4, which uses the energy difference between the subjets and the jet p_T to estimate $\cos\theta$ in the rest frame. Panel (b) shows the distribution of the cosine of the angle between the two subjets in the vector boson rest frame, derived through a Lorentz transformation as described in Equation A.10. Panels (c) and (d) display the distributions of the simple lab frame estimate of $\cos\theta$ in the transverse plane (Eq. A.5) for the primary and secondary subjets, respectively. Lastly, Panels (e) and (f) depict the full 3D angular calculation in the lab frame using Equation A.7, for both subjets. Although the engineered features individually offer limited discrimination between polarisation states, the full 3D angular calculation, particularly for the secondary subjet, shows promise when used in combination with other jet features.

4.2.2 Binary classification

The classification model results, encompassing a BDT and DNN, are presented through predicted probabilities and performance metrics. This comparative analysis underscores the effectiveness of each model in differentiating between longitudinal and transverse polarisations. Due to computational and time constraints, only ten percent of the total MC20 data is used for this analysis.

The optimal configurations for the models utilised in this thesis are outlined as follows. The boosted decision tree model, designed for binary logistic regression, features a maximum depth of 15 and a learning rate of 0.01. To bolster robustness and mitigate overfitting, both subsampling and column sampling rates are established at 0.8. The BDT is trained for a maximum of 10,000 iterations, with early stopping implemented if no improvements are observed over 10 consecutive rounds. The model's performance is assessed using the Log-loss function.

In contrast, the optimal deep neural network model comprises three hidden layers, each with 512 neurons and employing the ReLU activation function. The output layer consists of a single node with a sigmoid activation function. The learning rate is also set at 0.01, paired with a regularisation strength of 0.001 to help prevent overfitting by penalising excessively large weights in the network. The training process utilises the Adam optimiser, recognised for its efficiency in managing sparse gradients and its adaptive learning rate features. Similar to the boosted decision tree, the DNN employs log loss as its loss function. However, training is limited to 100 epochs, with early stopping activated after 5 epochs, facilitating rapid convergence and efficient training cycles.

The predicted probabilities for the top-performing classifiers are depicted in Figure 27. The stacked distributions in panels (a) and (c) demonstrate moderate distinction between polarisations. This distribution illustrates that both states are very similar, highlighting the difficulty of the problem. The normalised distributions in panels (b) and (d) further emphasise the polarisation proportions, adjusted to their respective counts. This shows that some longitudinal states can be classified with an underlying number of transverse states.

To further evaluate the performance of the BDT and DNN, several metrics are utilised, including ROC and Precision-Recall curves. The ROC curves shown in panels (a) and (c) of Figure 28 reveal areas under the curves (AUC) of 0.75 for the BDT and 0.60 for the DNN. These values indicate fairly low discriminatory power, suggesting that both models are inadequate for the classification task. Notably, the higher AUC for the BDT implies it still holds an advantage in distinguishing between polarisations.

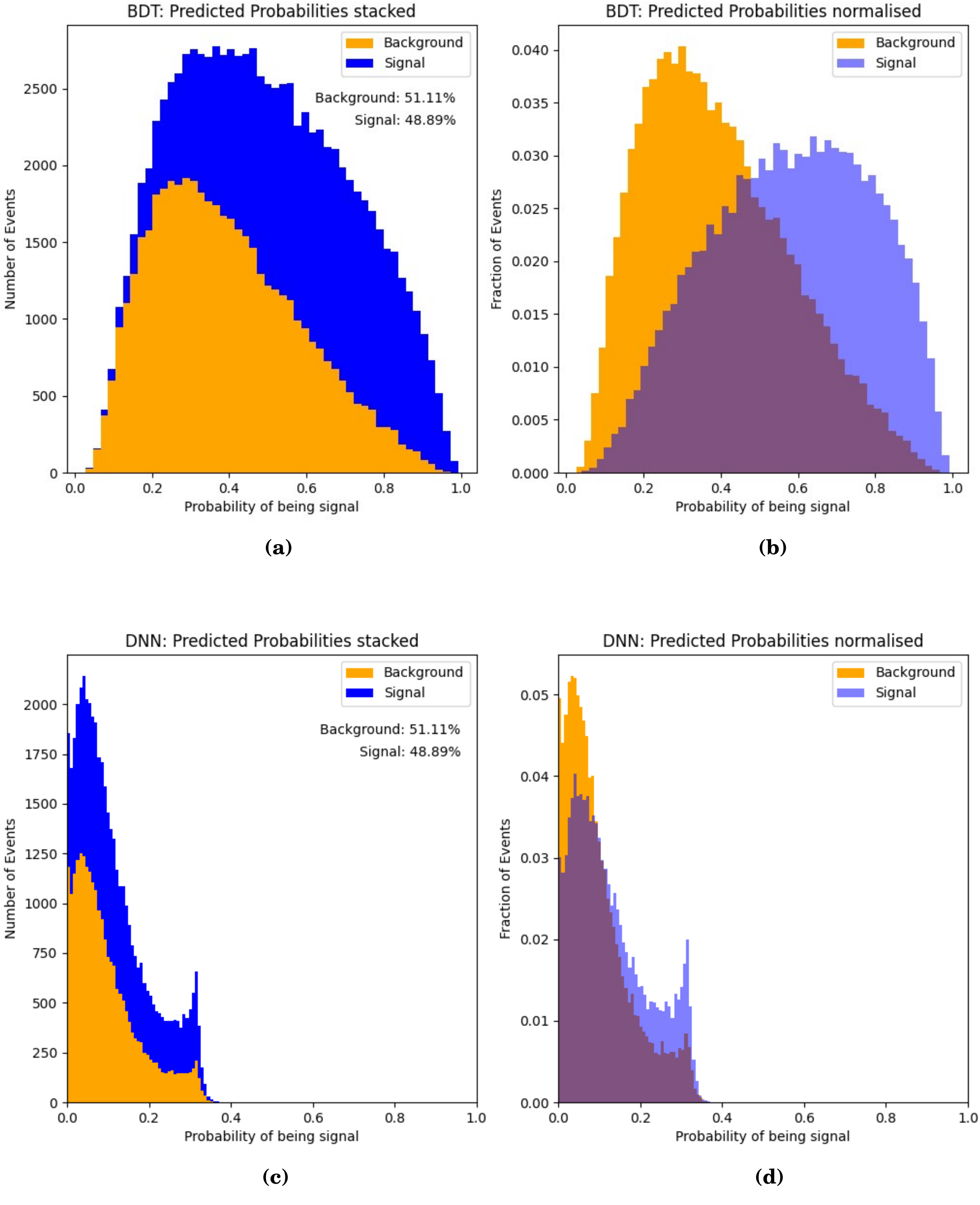


Figure 27: Comparison of predicted probabilities for the BDT and DNN classifiers. Panels (a) and (c) display stacked distributions of the predicted probabilities, with background events (orange) representing transverse polarisation states and signal events (blue) corresponding to longitudinal polarisations. The background accounts for 48.89% of the events, while the signal comprises 51.11%. Panels (b) and (d) show the normalised distributions, illustrating the fraction of events relative to the total within each class.

In addition, the Precision-Recall curves in panels (b) and (d) complement the ROC curves by illustrating the models' performance in terms of precision and recall across various probability thresholds. While the PR curve is particularly useful for imbalanced datasets, it also effectively reflects performance when classes are balanced, as is the case here. In such datasets, both the ROC and PR curves are likely to exhibit similar trends. This can be seen back in the plot, where the areas under the curves are 0.74 for the BDT and 0.58 for the DNN, values that align closely with those observed in the ROC analysis. When the proportions of positive and negative instances are comparable, changes in the false positive rate (used in the ROC curve) correlate more closely with changes in precision and recall metrics from the PR curve. This results in a more representative evaluation of overall performance. In summary, although both curves can yield comparable results in a balanced scenario, they emphasize different aspects of performance, making it beneficial to examine both for a comprehensive evaluation.

Supplementary performance plots for both the BDT and DNN models can be found in Appendix C: Polarisation Related Plots, providing additional validation and context for the results presented here.

Overall, the findings indicate that both the BDT and DNN models are suboptimal in their ability to classify polarisation states. A key observation during training is the significant gap between training error (≈ 0.15) and validation error (≈ 0.60), which notably exceeds the difference between training error and human-level error. This discrepancy indicates that the models are likely facing a variance problem, suggesting potential overfitting to the training data. To address this issue, several strategies can be employed, such as acquiring more data, implementing regularisation and dropout techniques, augmenting the dataset, and conducting hyperparameter searches. Since regularisation and dropout are already applied, and augmenting the data along with hyperparameter tuning requires careful planning and additional effort, the most straightforward approach is to increase the dataset size. This is particularly feasible, as only ten percent of the available data has been used thus far.

These results underscore the necessity for further investigation to achieve the desired performance levels, while also highlighting that the most accessible improvement lies in leveraging the additional data at hand.

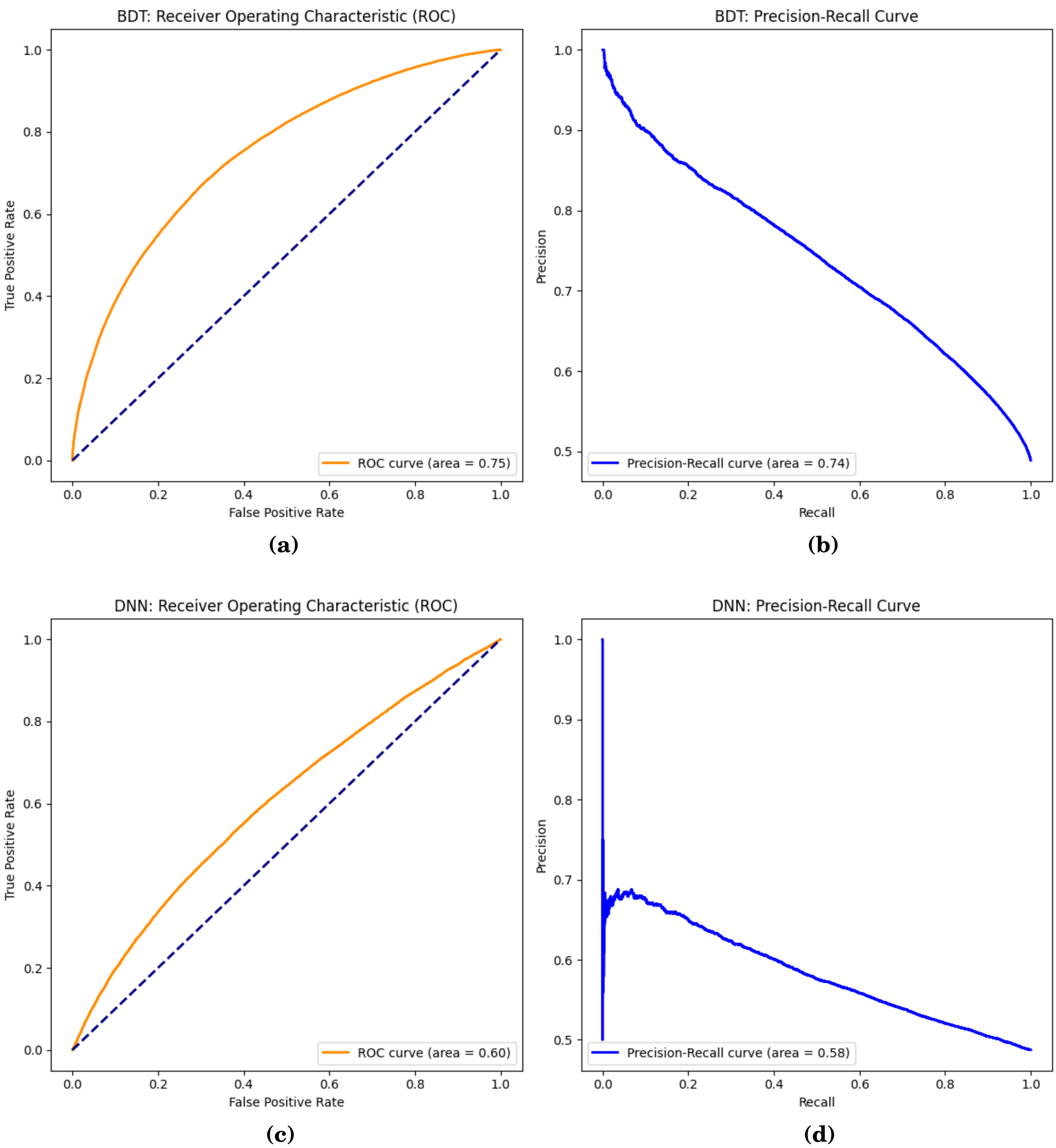


Figure 28: Comparison of performance metrics for the boosted decision tree (BDT) and deep neural network (DNN) models. The upper (a) and lower left (c) plots display the Receiver Operating Characteristic (ROC) curves, with areas under the curve (AUC) of 0.75 for BDT and 0.60 for DNN, indicating moderate discriminatory power. The upper right (b) and lower right (d) plots present the Precision-Recall curves, with AUC values of 0.74 for BDT and 0.58 for DNN, reflecting the models' limited ability to differentiate between signal and background events across varying probability thresholds.

4.2.3 Cosθ regression

The final results of this thesis consist of $\cos\theta$ regression performance plots. The model is assessed on how accurately it can estimate the parton-level $\cos\theta$ value. While various estimations of this angle have been tried above, none have shown discrimination power, therefore a more refined approach is done using a machine learning regression model.

Figures 29 and 30 illustrate the model's ability to estimate the $|\cos\theta|$ values for both W and Z bosons. Figure 29 presents the deviations between true and predicted values, centred around zero with a slight bias and a standard deviation of 0.17, reflecting overall consistency while highlighting room for improvement. Meanwhile, Figure 30 shows that the model performs reasonably well for larger $|\cos\theta|$ values, with predictions closely matching the target values. However, performance significantly declines near $|\cos\theta| = 0$, where predictions exhibit a greater spread, indicating reduced accuracy. As depicted in Figure 31, the model fails to predict values around zero, instead concentrating predictions near 0.1 and 0.3.

A likely cause for this discrepancy at low cosine values is the differing $\cos\theta$ distributions for the W and Z bosons, which exhibit the largest deviations at $|\cos\theta| = 0$ and $|\cos\theta| = 1$ (Fig. 6). To address this issue, the distributions for W and Z are trained separately, as depicted in Figure 32. While these separate distributions show slight improvements in the regression of the angle, the same issue persists around the small $|\cos\theta|$ values, with the model not predicting any values near zero. Further research is required to understand the underlying causes of this performance limitation.

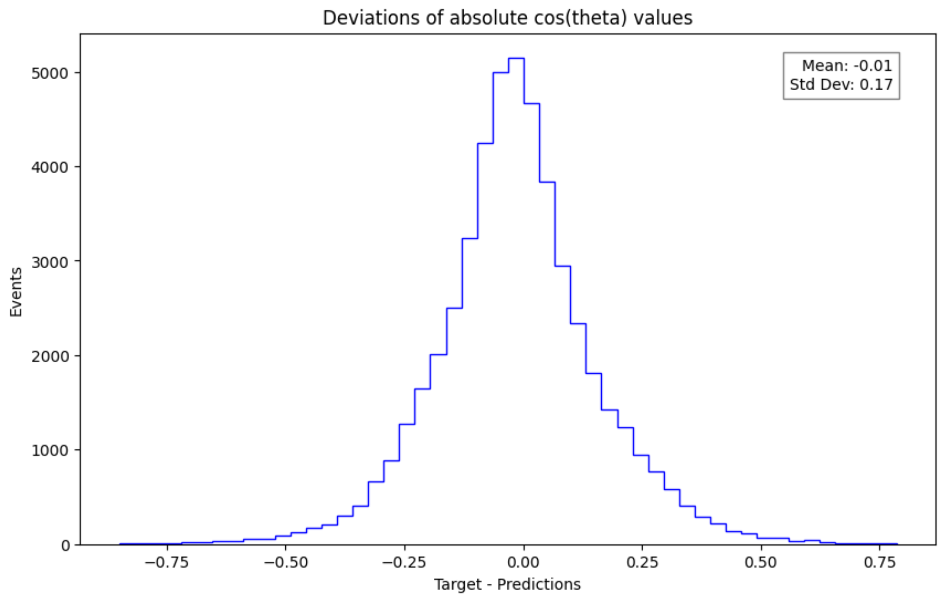


Figure 29: Histogram showing the deviations between true and predicted values of $\cos\theta$ for the regression model applied to W and Z bosons. The graph is centred at -0.01, with a standard deviation of 0.17.

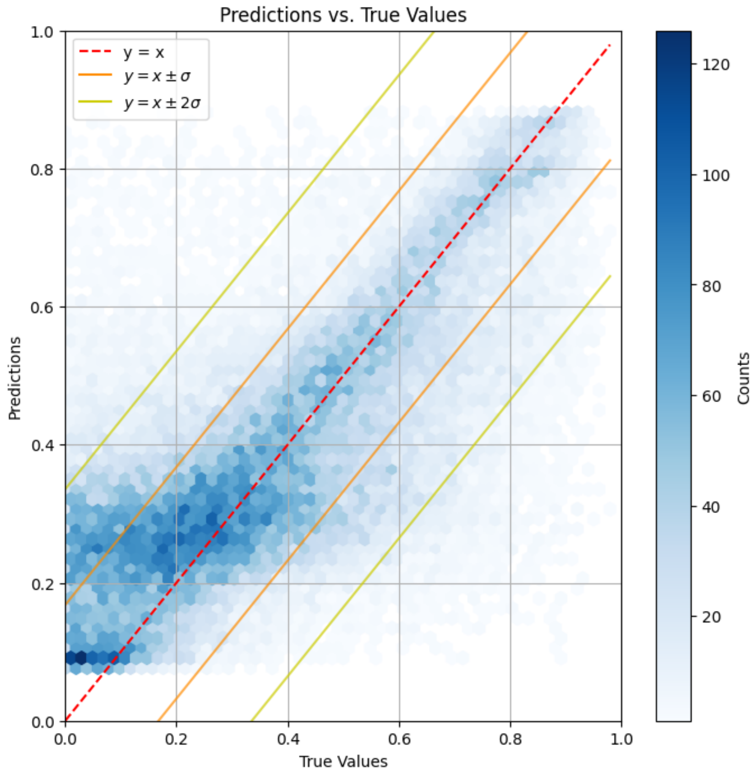


Figure 30: Predicted vs. true values for the regression model output, covering both transverse and longitudinal polarisations. The plot illustrates the model's performance for predicting $|\cos\theta|$ values of both W and Z bosons. The diagonal red dashed line represents the ideal case where predictions perfectly match the true values (y=x), while the yellow and orange lines denote deviations of $\pm1\sigma$ and $\pm2\sigma$, respectively. The density of the points, shown by the colour scale on the right, indicates where the model predictions are concentrated. The model performs reasonably well for larger $|\cos\theta|$ values, with predictions closely following the y=x line. However, the performance is noticeably worse near $|\cos\theta| = 0$.

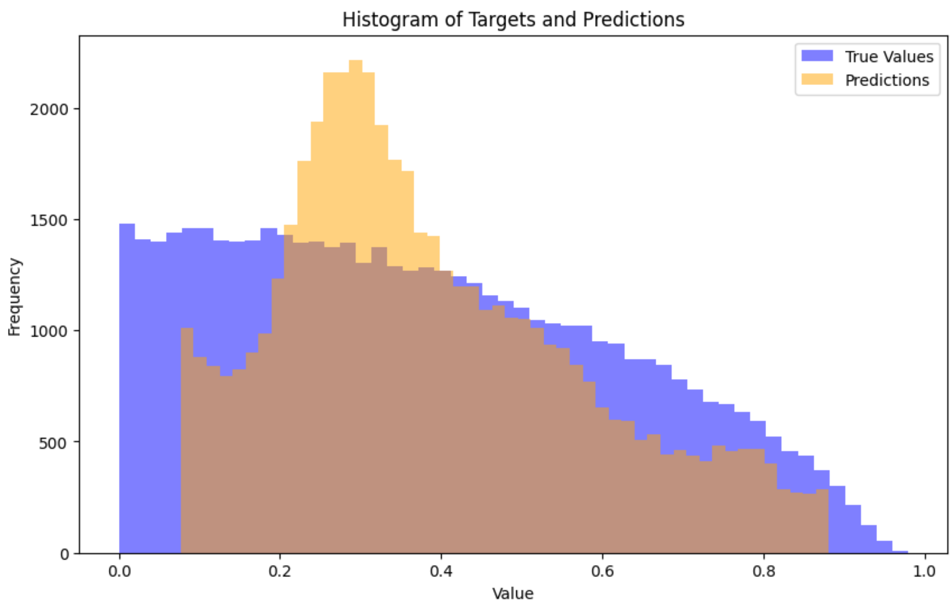
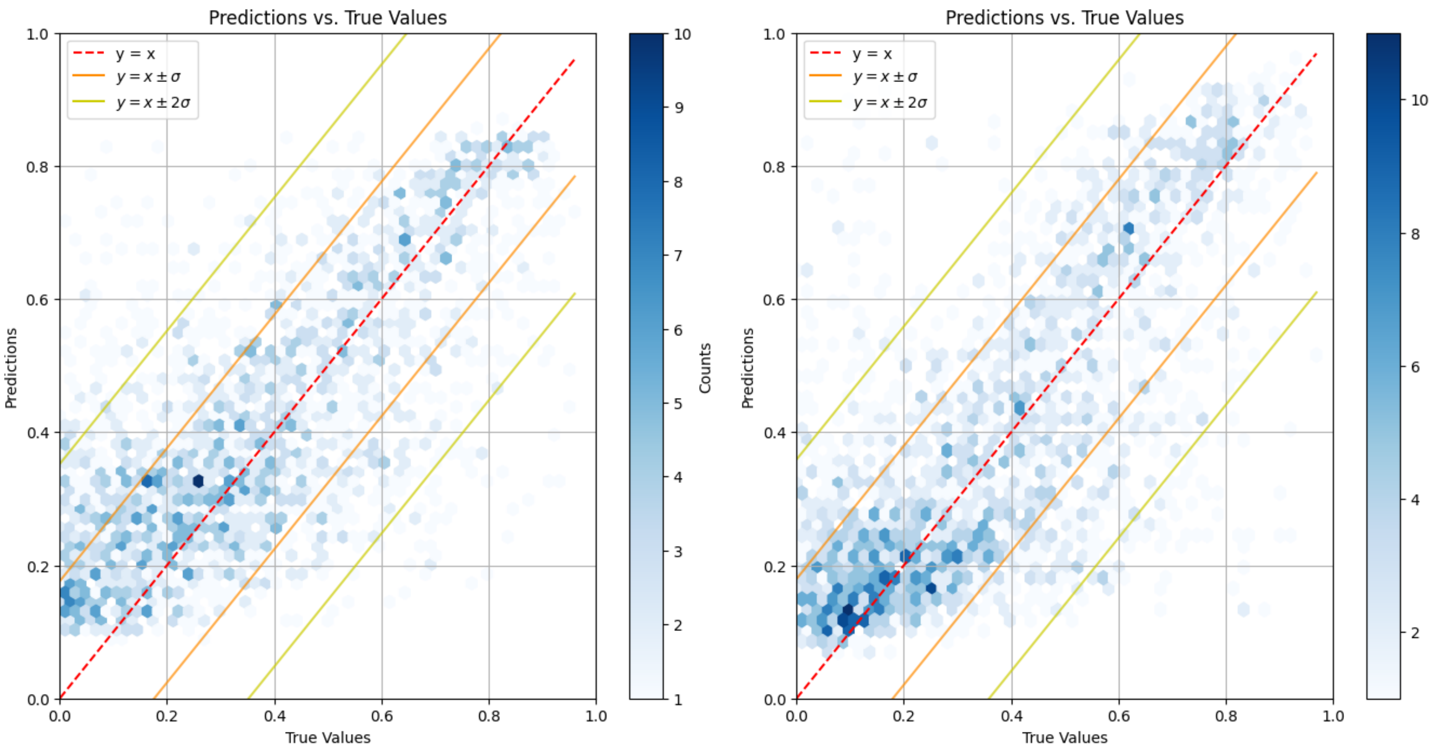


Figure 31: Histogram illustrating the distributions of true and predicted $\cos\theta$ values for the regression model applied to both W and Z bosons. The plot highlights the model's challenge in predicting lower $\cos\theta$ values, as it fails to generate predictions near zero, instead forming a peak around 0.3.



(a) **(b)**

Figure 32: Predicted vs. true values for the regression model output, covering both transverse and longitudinal polarisations of W and Z bosons separately. Plot (a) shows the model's performance for W boson $\cos\theta$ values, while plot (b) focuses on Z boson $\cos\theta$ values. The red dashed line represents the ideal case (y=x), with the yellow and orange lines indicating $\pm 1\sigma$ and $\pm 2\sigma$ deviations, respectively. The colour scale reflects the density of predictions, revealing strong model performance at higher $|\cos\theta|$ values, where predictions align with the y=x line. However, closer to $|\cos\theta| = 0$, the spread increases, indicating reduced accuracy.

5 Conclusion

This study presents key advancements in the classification of signal versus background events within the fully hadronic VVjj channel. Through detailed feature analysis and the application of ML models, the research identifies the most discriminative features, demonstrating the effectiveness of both boosted decision trees and deep neural networks for classification tasks.

Features such as the DisCo score, jet mass, number of tracks and D2 score prove critical in distinguishing between signal and background events, with minimal contribution from secondary jet features. The analysis shows consistent performance across various signal models and mass ranges, highlighting the robustness of these features. As the resonance mass increases, dijet features related to energy and mass, along with jet energy and p_T, become more discriminative, highlighting their increasing importance at higher masses.

Both the BDT and DNN models exhibit strong discriminatory power, with ROC AUC scores of 0.93 and 0.90, respectively. While the BDT slightly outperforms the DNN in classification accuracy, the DNN's ability to capture complex data patterns underscores its potential for future applications. Precision-Recall metrics further affirm the models' reliability in classifying signal events while minimising false positives, crucial in high-energy physics analyses.

In summary, the integration of feature analysis and machine learning techniques creates a robust classification framework that paves the way for future studies.

Despite these efforts, polarisation classification has not yet achieved strong results. Discrimination between the various features remains limited, apart from some JSS variables that show more noticeable differences between polarisations. However, the models have not managed to extract significantly more discriminating information, resulting in ROC AUC scores of around 0.75 at best. This highlights the need for innovative approaches to improve performance.

One promising avenue is decay angle regression, which offers a model-independent intermediate result. This angle can be used as an additional feature to enhance existing classifiers or potentially as a standalone tool. Moreover, it provides valuable opportunities for experimental calibration. Initial results from this thesis suggest that the decay angle can be estimated within the 1σ range and performs reasonably well at higher $\cos\theta$ values. However, predictions for lower $\cos\theta$ values, especially around zero, remain more scattered. The underlying cause of this discrepancy has not been identified in this project, but it does not appear to stem from the differing angular distributions of the W and Z bosons.

Overall, this research establishes a strong foundation for future efforts, with decay angle regression and continued model refinement offering promising avenues for improving polarisation tagging in the future.

6 Discussion

The signal vs. background classification for VVjj in this study has demonstrated excellent performance, suggesting diminishing returns from further exploration with similar models. While larger or alternative models could be explored, they are unlikely to yield substantial improvements at this stage.

In contrast, the challenges associated with polarisation tagging are more pronounced. The current binary classification of transverse and longitudinal polarisations falls short of satisfactory performance and requires substantial improvement. Developing and testing more sophisticated models could significantly enhance classification efficiency. With sufficient computational power and resources, incorporating all constituent data into the classifier is anticipated to enhance its discrimination capabilities, particularly considering the identified variance problem. Nevertheless, the exact degree of improvement remains uncertain.

Moreover, insights from existing methods, such as the neural networks developed for hadronic tau polarisation tagging, may offer valuable guidance for vector boson polarisation tagging. However, adapting these techniques is far from straightforward. A promising alternative lies in using reclustered information from methods like LundPlane or AntiKt, which could provide improvements over the track-reconstructed data that has been the primary focus of this thesis.

In addition to enhancing the classifier, exploring $\cos\theta$ regression offers a promising avenue for further research. By incorporating this regression as an additional feature in the classifier, greater discrimination between polarisation types could be achieved. Furthermore, integrating $\cos\theta$ into an auxiliary task may not only enhance polarisation classification but also broaden its applicability in other contexts. To advance this approach, it is essential to add $\cos\theta$ information to the MC20 data and employ a regression model to estimate the truth-level values. This would enable a detailed evaluation of how incorporating $\cos\theta$ affects classification performance and test its efficiency in distinguishing between transverse and longitudinal polarisations. Such an approach could even prompt a reconsideration of the necessity of classification altogether.

A major challenge remains the lack of precise theoretical predictions at present, which are crucial for experimental measurements. Despite significant efforts by the theoretical community, more accurate predictions, particularly beyond leading order (LO), are still required. This raises concerns about the quality of the simulated polarisation samples used in this study. While LO MadGraph samples offer a solid foundation, further investigation into higher-order corrections, including QCD effects and interference contributions, is essential to assess their impact on polarisation classification. Tree-level simulations, in combination with parton shower modelling, may offer sufficient precision for training purposes, but this approach needs careful evaluation.

A final consideration is how to define a well-performing polarisation tagger. While clean simulated samples may exhibit strong performance, real-world complications such as pileup, background noise, and detector resolution must be adequately addressed. Furthermore, given that polarisation is not p_T-independent, the same p_T cut can affect different polarisation states in distinct ways. Additionally, a truly effective tagger needs to be sensitive to calibration, as uncalibrated inputs pose a significant risk to accurate classification. Experimental calibration approaches may be particularly promising in ttbar and V+jets events, where W polarisation fractions predominantly manifest as longitudinal or transverse. However, no concrete solutions have emerged to tackle this critical issue.

7 Outlook

The key takeaway from this thesis presents a promising outlook for future work: polarisation tagging is an emerging field with great potential for growth. While substantial progress has been made, it remains in its early stages, requiring considerable effort before it can be fully integrated into particle physics analyses and the ATLAS framework.

Moving forward, several crucial questions must be addressed. Firstly, is it necessary to classify transverse and longitudinal polarisations at the jet level, or would alternative strategies, such as measuring polarisation ratios, prove more effective? Additionally, determining whether classification should focus on individual bosons or entire events poses further challenges, especially when applying these methods across various analyses. Furthermore, the classification approach may benefit from incorporating p_T and η dependence to enhance its applicability in different studies. Lastly, it is important to evaluate whether fat-jet substructure is essential for effective polarisation tagging or if other global observables might prove more sensitive to polarisation signals.

As the field continues to evolve, embracing innovative methods and perspectives will be essential to unlock the full potential of polarisation tagging, ultimately contributing to a deeper understanding of particle physics.

References

[1] Glenn T. Seaborg. „Evolution of the modern periodic table". In: *Journal of the Chemical Society, Dalton Transactions.* (1996). p. 3899–3907.

[2] David Griffiths. *Introduction to elementary particles*. 2nd ed. (John Wiley & Sons, 2020).

[3] Peter Steinberg and ATLAS Collaboration*, et al*. „Recent heavy-ion results with the AT-

LAS detector at the LHC". *Journal of Physics G: Nuclear and Particle Physics* **38** (2011). p. 124004.

[4] Mark Thomson. *Modern particle physics* (Cambridge University Press, 2013).

[5] *Standard Model Image,* URL: `https://en.wikipedia.org/wiki/Standard_Model` (accessed: Jun. 2024).

[6] Bryce S. DeWitt. „Quantum theory of gravity. I. The canonical theory". *Physical Review* **160** (1967). p. 1113.

[7] Laurent Canetti, Marco Drewes and Mikhail Shaposhnikov. „Matter and Antimatter in the Universe". In: *New Journal of Physics* **14** (2012). p. 095012.

[8] Gianfranco Bertone and Dan Hooper. „History of dark matter". In: *Reviews of Modern Physics* **90** (2018). p. 045002.

[9] Katherine Garrett and Gintaras Dūda. „Dark matter: A primer". In: *Advances in Astronomy* **2011** (2011). p. 968283.

[10] Murray Gell-Mann, Pierre Ramond and Richard Slansky. *Murray Gell-Mann: Selected Papers* (World Scientific, 2010). p. 266–272.

[11] Peter Fisher, Boris Kayser and Kevin S. McFarland. „Neutrino mass and oscillation". *Annual Review of Nuclear and Particle Science* **49** (1999). p. 481–527.

[12] Lisa Randall and Raman Sundrum. „Large mass hierarchy from a small extra dimension". In: *Physical review letters* **83** (1999). p. 3370.

[13] Bonifatius Paulus Josephus Stienen. „Working in high-dimensional parameter spaces: Applications of machine learning in particle physics phenomenology". Ph.D. thesis. SI:[Sn] (2021).

[14] ATLAS Collaboration. „Search for diboson resonances with boson-tagged jets in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector". In: *Physics Letters B* **777** (2018). p. 91–113.

[15] ATLAS Collaboration. „Search for diboson resonances in hadronic final states in 139 fb^{-1} of pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector". In: *Journal of High Energy Physics* **2019** (2019).

[16] Johannes Erdmann. „Experimental results in the boosted regime". *Tech. rep.*. ATL-COM-PHYS-2014-1461 (2014).

[17] Guido Altarelli, Barbara Melé and M Ruiz-Altaba. „Searching for new heavy vector bosons in p colliders". In: *Zeitschrift für Physik C Particles and Fields* **45** (1989). p. 109–121.

[18] Duccio Pappadopulo, *et al*. „Heavy vector triplets: bridging theory and data". In: *Journal of High Energy Physics* **2014** (2014). p. 1–50.

[19] Carsten Bittrich. „Study of Polarization Fractions in the Scattering of Massive Gauge Bosons W±Z→ W±Z with the ATLAS Detector at the Large Hadron Collider". Ph.D. thesis. Technische Universität Dresden (Apr. 2015).

[20] Songshaptak De, Vikram Rentala and William Shepherd. „Measuring the polarization of boosted, hadronic W bosons with jet substructure observables". *arXiv preprint, arXiv:2008.04318* (2020).

[21] Sofia Adorni Braccesi Chiassi. „Vector Boson Tagging in the Context of the Search for New Physics in the Fully Hadronic Final State". Ph.D. thesis. Geneva U. (2021).

[22] Y. Gao. „Normalised differential distributions of the cross section over the cosine of the decay angle of a hypothetical new particle X". *Minutes of jet tagging meeting* (2 Nov. 2023). CERN.

[23] *ROC Curve Image,* URL: `https://en.wikipedia.org/wiki/File:Roc_curve.svg` (accessed: Sept. 2024).

[24] Hong Hui Tan and King Hann Lim. *2019 7th international conference on smart computing & communications (ICSCC)* (IEEE, 2019). p. 1–4.

[25] Takuya Nobe. „Jet tagging and scale factors: Introduction to jet tagging". In: *Proceedings of the ATLAS Hadronic Calibration Workshop* (30 Aug. 2014). Ottawa, Canada.

[26] Huilin Qu, *et al*. „Weaver". `https://github.com/hqucms/weaver-core` (2024). Version 0.4.17, Accessed: Aug. 2024.

[27] Shudong Wang. „Preliminary Result on W Jet Tagger with ParticleNet". *Minutes of IHEP ML Innovation Meeting* (2022). IHEP, Beijing, China.

[28] Gregor Kasieczka and David Shih. „Robust jet classifiers through distance correlation". *Physical review letters* **125** (2020). p. 122001.

[29] Andrew J Larkoski, Ian Moult and Duff Neill. „Analytic boosted boson discrimination". *Journal of High Energy Physics* **2016** (2016). p. 1–133.

Acknowledgements

This thesis owes much to the support and collaboration of many wonderful people. I am deeply grateful for the friendships and connections made during my project at Nikhef and for the amazing experiences throughout the year.

First and foremost, I want to express my heartfelt appreciation to Flavia de Almeida Dias for overseeing my master's project at Nikhef. Her constant support and valuable advice were pivotal to the success of this work. Flavia was always accessible and ready to address any questions, providing insightful feedback during our meetings. I am particularly thankful for the freedom she gave me to explore the topic in my own way, allowing me to follow the approach that I found most interesting. I also appreciate the opportunities she facilitated, including my visits to CERN and Toulouse, which greatly enriched my project and expanded both my knowledge and my network.

I am sincerely grateful to Dylan van Arneman for his support as my daily supervisor. His insightful suggestions and follow-up questions were invaluable, and his willingness to assist, whether in person or over text, made a significant difference.

My appreciation also extends to Wouter Verkerke for serving as my second examiner and for fostering a positive social environment within the ATLAS group. I am thankful to the entire Nikhef ATLAS group for contributing to this welcoming atmosphere. Thank you all for the engaging conversations outside our weekly meetings and for making my time at Nikhef so enjoyable. I will always cherish the ping-pong games we had!

A special thank you goes to Victoria, my wonderful girlfriend, for her unwavering support throughout this journey. I look forward to finally closing the distance as I embark on my next adventure as a PhD student in Austria. Lastly, I am profoundly grateful to my parents, Klaas and Cora, for their support and encouragement throughout my master's project, especially during the most challenging times. I would like to dedicate this work to my mother, wishing she could have been here to witness it.

I would like to extend my thanks to CERN for the successful operation of the LHC and to the support staff from our institutions, whose efforts ensure the efficient operation of ATLAS. Additionally, I am grateful to the COMETA group for providing a platform and hosting multiple workshops that advanced my research on polarisation and multi-boson machine learning.

Finally, I acknowledge the use of AI as a valuable tool in developing this thesis, assisting with grammar, spelling, coding support, and generating the image in the popular scientific summary. That said, it is important to clarify that AI was only used as a supportive tool, providing feedback and suggestions rather than producing content independently. All the work presented in this thesis is my own, created through my efforts, unless otherwise cited.

Appendices

The appendices of this thesis present supplementary materials generated throughout the project. Each appendix is organised to focus on a specific aspect of the analysis:

- Appendix A: Cosθ approximation calculated from the dot product.
- Appendix B: Additional signal vs. background plots, including less significant feature analysis plots and extra machine learning validation metrics.
- Appendix C: Additional polarisation-related plots, featuring secondary feature analysis, extra machine learning validation metrics, and $\cos\theta$ regression plots.

The scripts used for the analysis can be found on the GitLab page: `https://gitlab.cern.ch/dsloot/polarisation-tagging`

Appendix A: Cosθ dot product estimations

This appendix outlines the derivation of the estimations for the cosine of the angle between two vectors based on the dot product, expressed as:

$$\vec{a}\cdot\vec{b} = |\vec{a}||\vec{b}|\cos\theta. \tag{A.1}$$

This can be reformulated to yield the cosine of the angle between the two vectors:

$$\cos\theta = \frac{\vec{p}_1\cdot\vec{p}_2}{|\vec{p}_1||\vec{p}_2|}, \tag{A.2}$$

where $\vec{p} = (p_x, p_y, p_z)$, with $p_x = p_T\cos\phi$, $p_y = p_T\sin\phi$, and $p_z = p_T\sinh\eta$. The momentum components of the jet and the subjet are utilised to calculate the angle between the vector boson and its decay products. By limiting the analysis to the lab frame and focusing solely on the transverse components p_x and p_y, the expression simplifies to:

$$\cos\theta = \frac{p_x^{\text{jet}} p_x^{\text{subjet}} + p_y^{\text{jet}} p_y^{\text{subjet}}}{\sqrt{(p_x^{\text{jet}})^2 + (p_y^{\text{jet}})^2}\sqrt{(p_x^{\text{subjet}})^2 + (p_y^{\text{subjet}})^2}}. \tag{A.3}$$

Filling in the x and y components gives

$$\cos\theta = \frac{p_T^{\text{jet}} p_T^{\text{subjet}}(\cos\phi^{\text{jet}}\cos\phi^{\text{subjet}} + \sin\phi^{\text{jet}}\sin\phi^{\text{subjet}})}{\sqrt{(p_T^{\text{jet}}\cos\phi^{\text{jet}})^2 + (p_T^{\text{jet}}\sin\phi^{\text{jet}})^2}\sqrt{(p_T^{\text{subjet}}\cos\phi^{\text{subjet}})^2 + (p_T^{\text{subjet}}\sin\phi^{\text{subjet}})^2}}. \tag{A.4}$$

Utilising the identities $\cos x\sin x + \cos y\sin y = \cos(x-y)$ and $\cos^2 x + \sin^2 x = 1$, this can be expressed as:

$$\cos\theta = \frac{p_T^{\text{jet}} p_T^{\text{subjet}} \cos\Delta\phi}{p_T^{\text{jet}} p_T^{\text{subjet}}} = \cos\Delta\phi, \tag{A.5}$$

indicating that the decay angle is directly related to the difference in azimuthal angle.

The second approximation for $\cos\theta$ incorporates the full three-dimensional orientation by including p_z terms in Equation A.3. This results in additional sinh terms, leading to:

$$\cos\theta = \frac{p_T^{\text{jet}} p_T^{\text{subjet}} (\cos\Delta\phi + \sinh\eta^{\text{jet}} \sinh\eta^{\text{subjet}})}{\sqrt{(p_T^{\text{jet}})^2 (1+\sinh^2\eta^{\text{jet}})}\sqrt{(p_T^{\text{subjet}})^2 (1+\sinh^2\eta^{\text{subjet}})}}. \tag{A.6}$$

Using the identity $1 + \sinh^2 = \cosh^2$ gives:

$$\cos\theta = \frac{\cos\Delta\phi + \sinh\eta^{\text{jet}} \sinh\eta^{\text{subjet}}}{\cosh\eta^{\text{jet}} \cosh\eta^{\text{subjet}}}. \tag{A.7}$$

These decay angle approximations using the dot product are expected to provide some discrimination power. However, these estimations are conducted within the lab frame and to obtain the true distributions for the decay angle of different polarisations, a transformation to the rest frame of the vector bosons is necessary. This requires a calculated boost factor to provide an accurate estimate of $\cos\theta$. Upon transformation to the vector bosons' rest-frame, the components of $\vec{p}^{\text{jet}}$ become zero, making it impossible to retrieve the decay angle between the vector boson and its decay products. Therefore, an alternative approach is employed, estimating the angle between the two decay products in the vector boson's rest-frame.

Starting from Equation A.2, where $\vec{p}_1$ and $\vec{p}_2$ are now given by the momentum vectors of the two subjets. For the Lorentz transformation, the boost factor is computed as

$$\beta = \frac{p_z^{\text{Jet}}}{E^{\text{Jet}}} \tag{A.8}$$

and

$$\gamma = \frac{1}{\sqrt{1-\beta^2}}. \tag{A.9}$$

The Lorentz transformation is then applied as follows:

$$\begin{aligned} p'_x &= p_x, \\ p'_y &= p_y, \\ p'_z &= \gamma(p_z - \beta E). \end{aligned} \tag{A.10}$$

Finally, the $\cos\theta$ formula (Eq. A.2) is filled in using these transformed components.

Appendix B: Signal vs. Background Plots

This appendix contains supplementary plots that provide deeper insights into the signal vs. background classification task. These figures focus on feature distributions for various signal models and resonance masses, alongside additional validation metrics for the machine learning models employed.

The analysis begins with an examination of the jet features for the Radion model signal at a resonance mass of 2000 GeV, compared to various background processes. The primary jet features, such as energy, transverse momentum, pseudorapidity, azimuthal angle, and the WZDNN score, are displayed in Figure 33. These features are instrumental in differentiating the characteristics of signal events from the background. Additionally, panel (f) shows the distribution of the number of Large-R jets, a factor critical to identifying high-mass resonances.

Figure 34 extends this analysis by comparing jet feature distributions across different signal models, specifically the Radion Model, Heavy Vector Triplet Model, and Bulk Randall-Sundrum Model. While most features are consistent, subtle variations in the WZDNN score reflect the different vector boson configurations (WW, WZ, ZZ) associated with each signature.

In Figure 35, the relationship between primary jet features, particularly energy and p_T, and the resonance mass in the Radion model is explored. As anticipated, these features increase proportionally with the resonance mass, illustrating the physical correlation between the resonance mass and vector boson decays. Interestingly, other features appear largely unaffected by changes in mass, as highlighted in subsequent plots.

The next section focuses on secondary jet features. Figure 36 presents the most discriminative secondary jet features for the Radion model signal at 2000 GeV in comparison to background processes. Key features, such as the DisCo score, jet mass, and the D2 score, demonstrate significant separation between signal and background, reaffirming their value in the classification task. Figure 37 supplements this with additional secondary features, demonstrating their further potential in distinguishing between signal and background.

To ensure consistency across different models, Figures 38 and 39 compare secondary jet features for the Radion, Heavy Vector Triplet, and Bulk Randall-Sundrum models. The majority of features remain consistent across the models, with minor differences in jet mass and WZDNN score reflecting the nuances in vector boson configurations for each signal.

Figures 40 and 41 shift attention to variations in resonance mass. These plots illustrate how secondary jet features, particularly the DisCo score and jet mass, behave across different resonance masses. While energy and p_T scale with the resonance mass, most secondary features exhibit stability, demonstrating that the classification remains robust across a range of resonance masses.

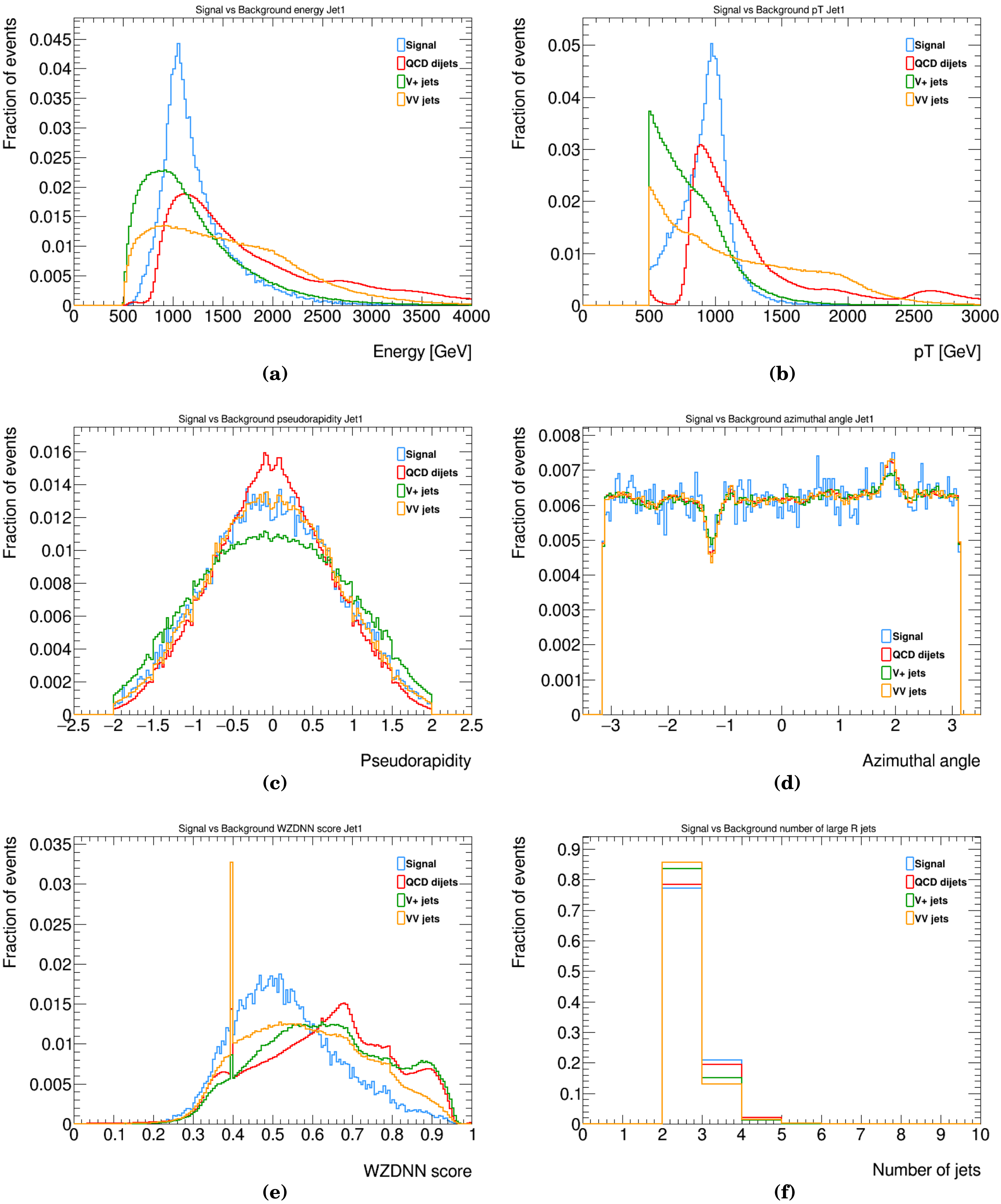
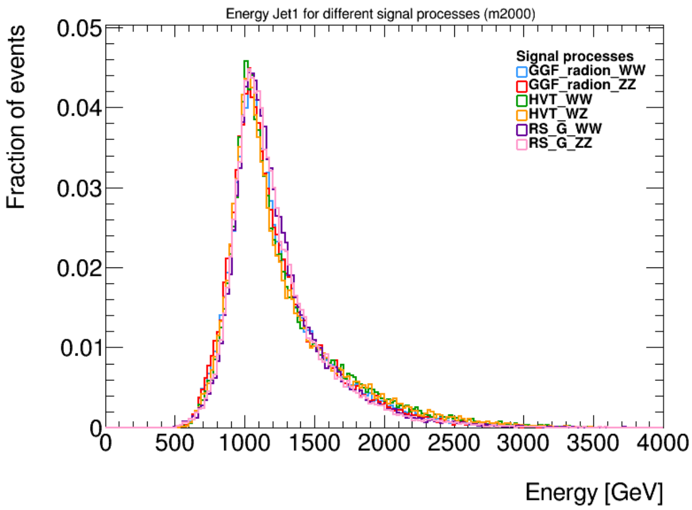
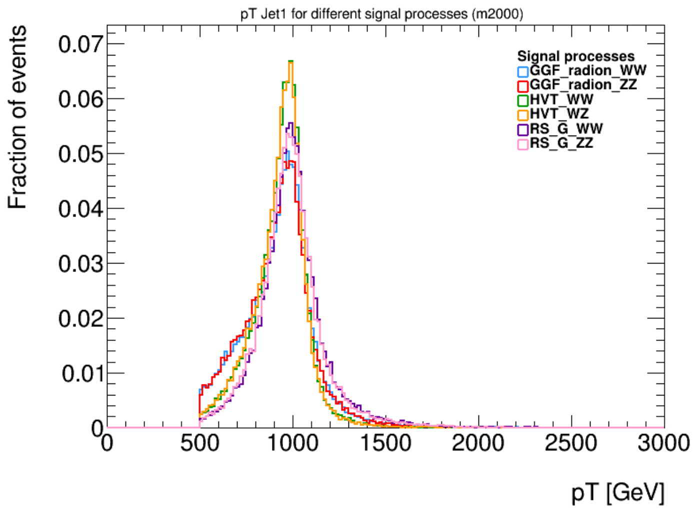


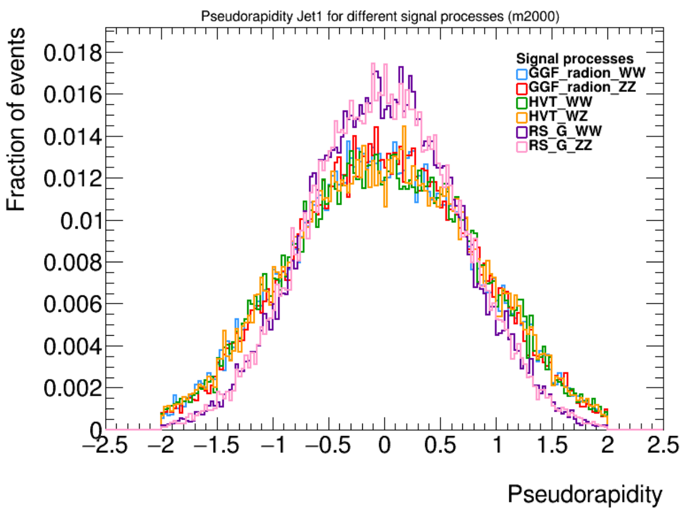
Figure 33: Distributions of additional jet features for the Radion model signal at m = 2000 GeV, compared to various background processes. Panels (a) to (e) display primary jet features, including energy, transverse momentum (p_T), pseudorapidity, azimuthal angle, and WZDNN score. Panel (f) presents the distribution of the number of Large-R jets.



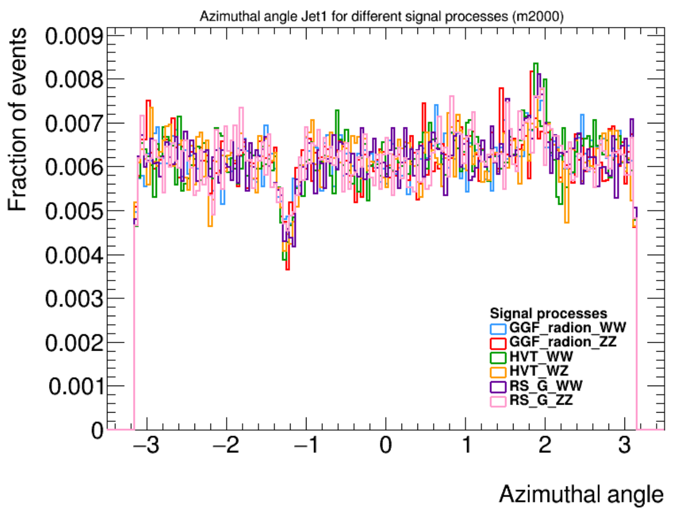
(a)



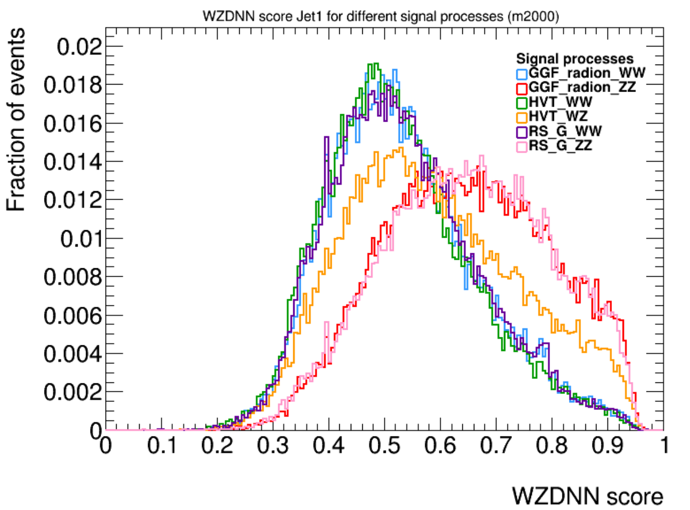
(b)



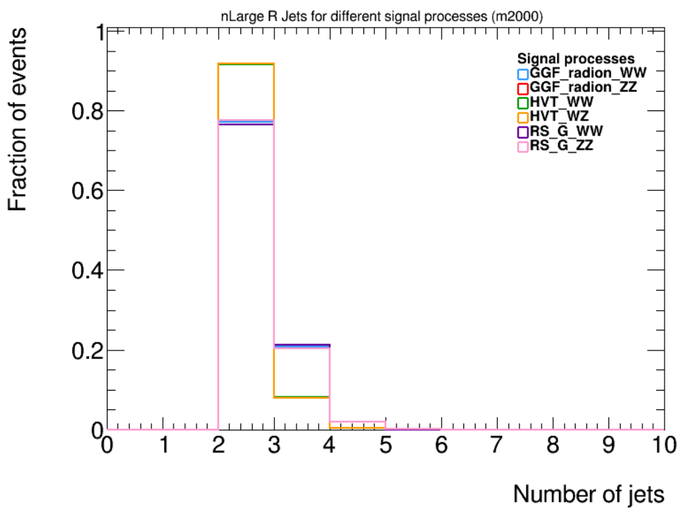
(c)



(d)



(e)



(f)

Figure 34: Distributions of additional jet features for various signal models, including the Radion Model, Heavy Vector Triplet Model, and Bulk Randall-Sundrum Model. Most features show consistency across these models, with minor variations in the WZDNN score reflecting the different vector boson configurations (WW, WZ, ZZ) associated with each signal.

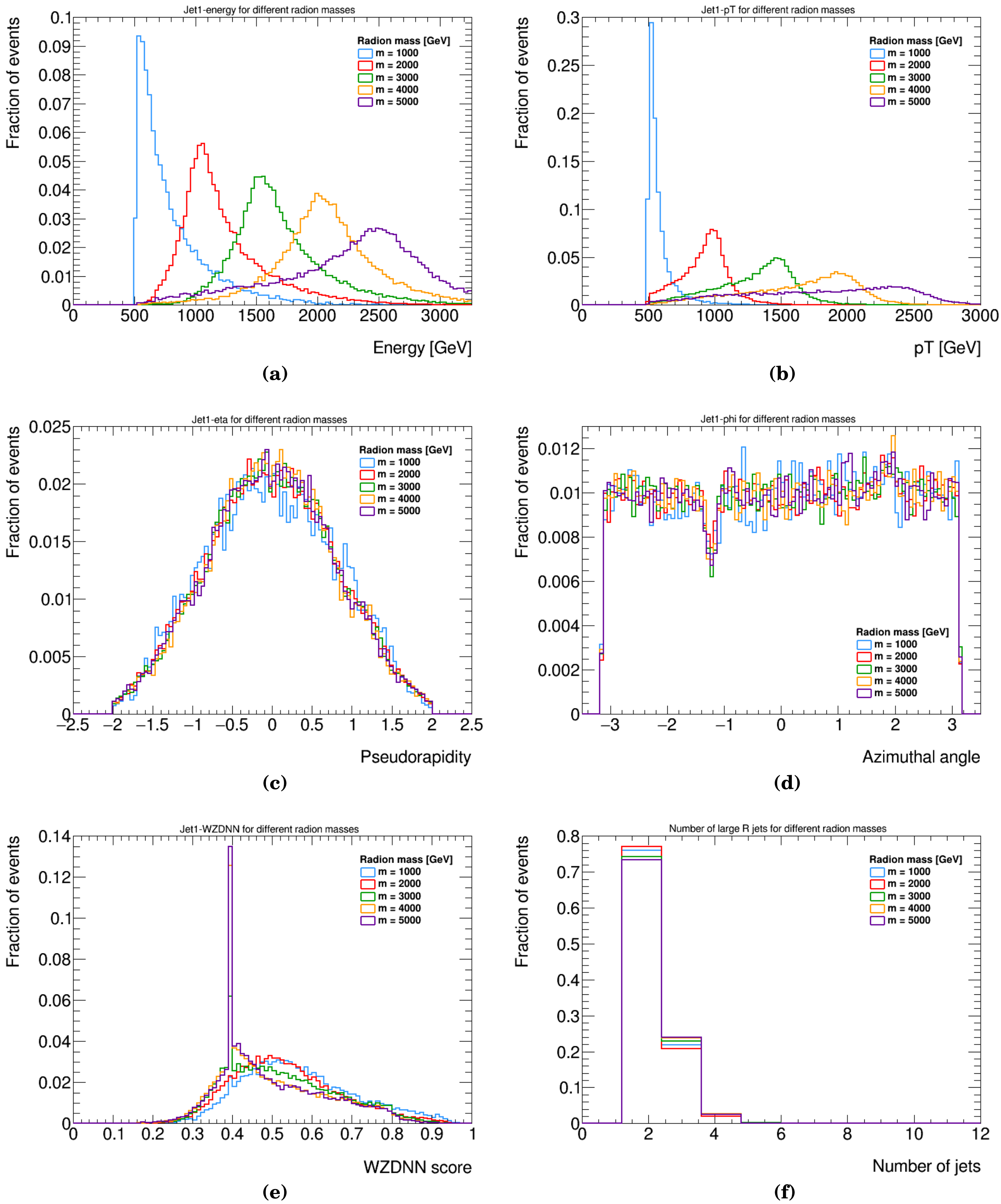


Figure 35: Distributions of additional features for the Radion model signal across various resonance masses. Panels (a) and (b) illustrate the primary jet features of energy and transverse momentum (p_T), which scale with resonance mass and reflect their connection to the vector bosons decaying from the heavy resonance. The other features remain consistent across different resonance masses.

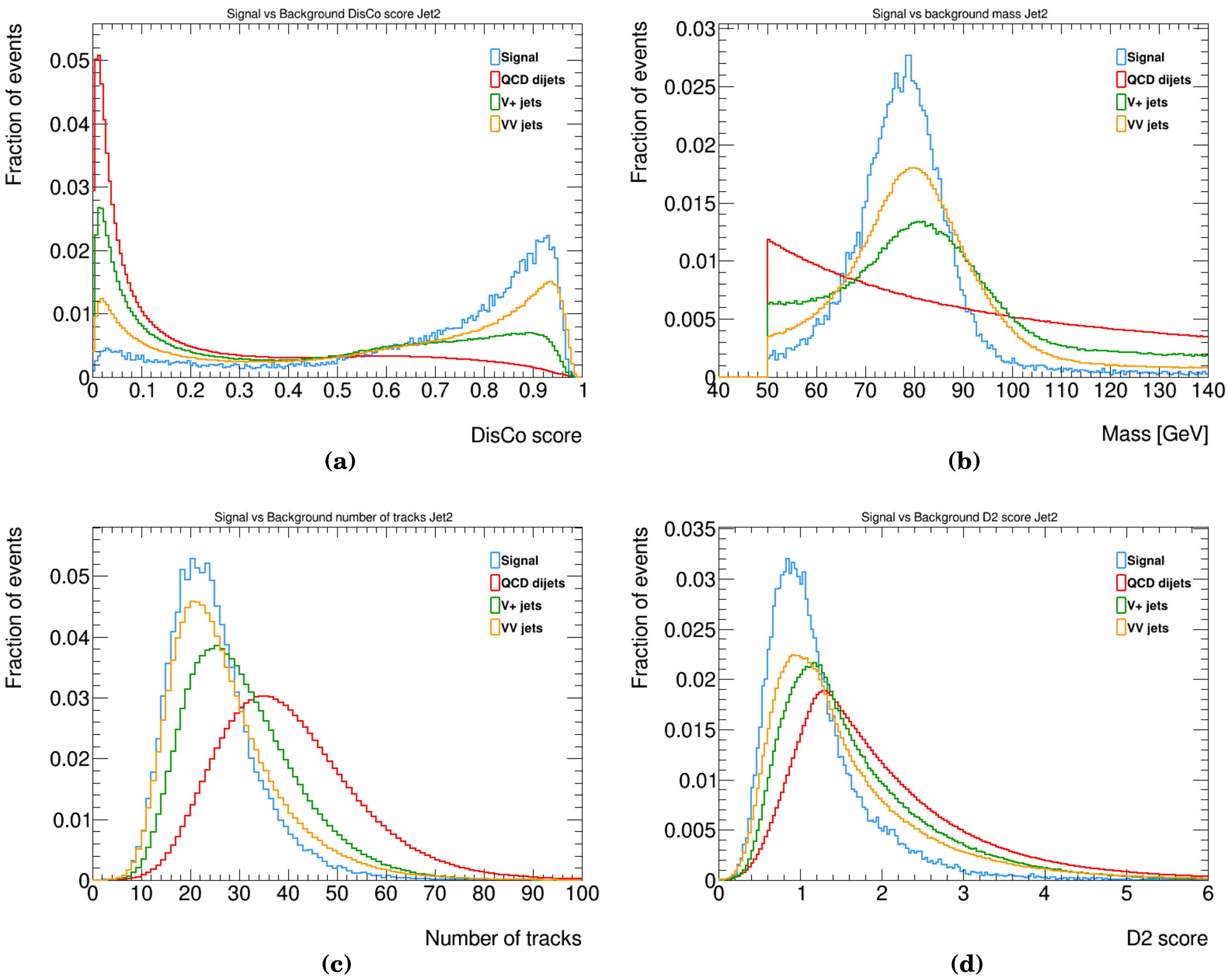
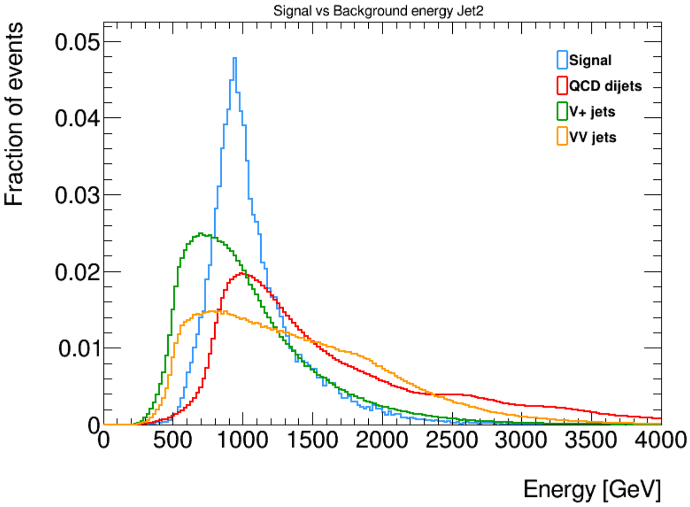
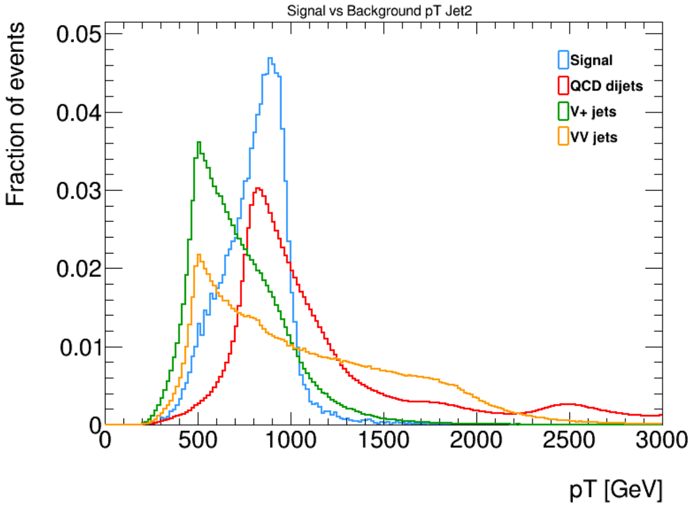


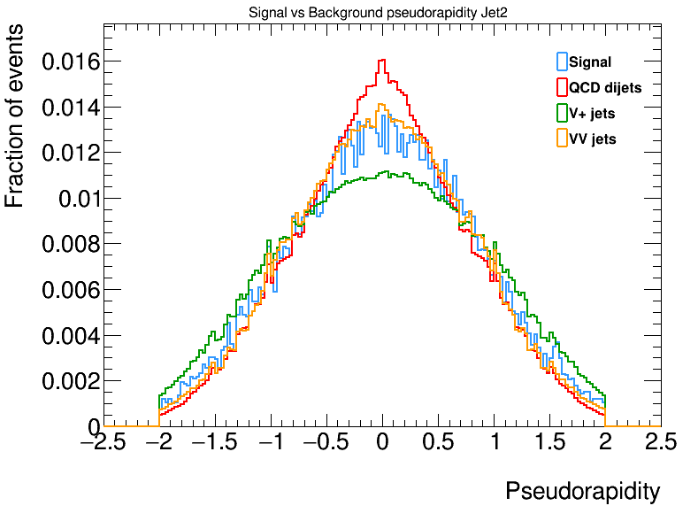
Figure 36: Distributions of the most discriminative secondary jet features for the Radion model signal (m = 2000 GeV) compared to various backgrounds. The features include the DisCo score, jet mass, number of tracks, and the D2 score.



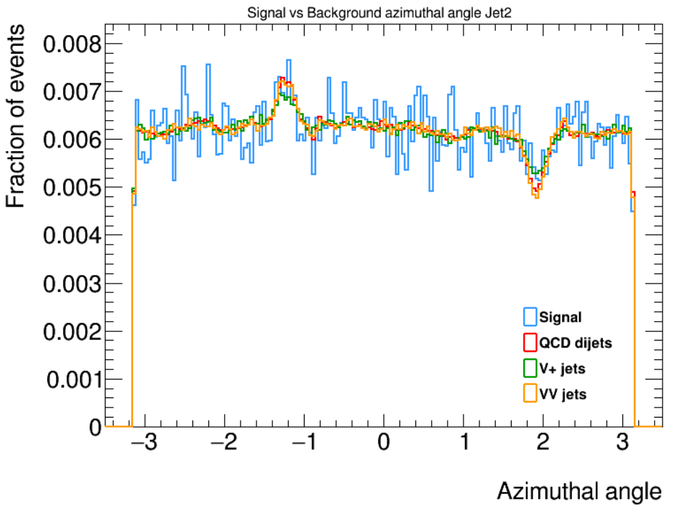
(a)



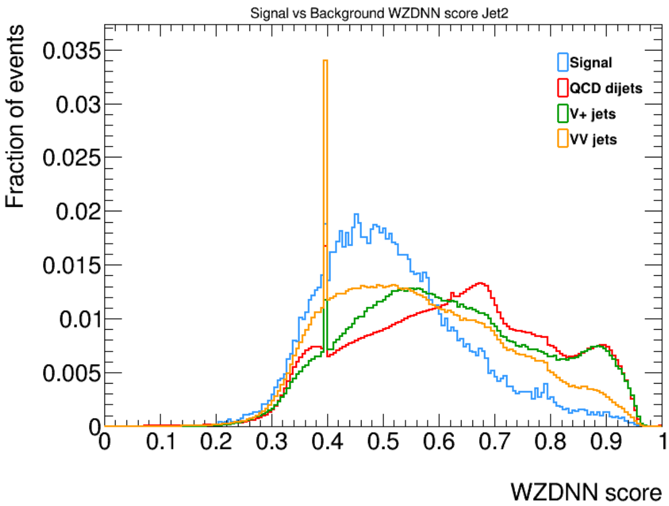
(b)



(c)



(d)



(e)

Figure 37: Distributions of additional secondary jet features for the Radion model signal at m = 2000 GeV, compared to various background processes. The features include jet energy, transverse momentum (p_T), pseudorapidity, azimuthal angle, and WZDNN score.

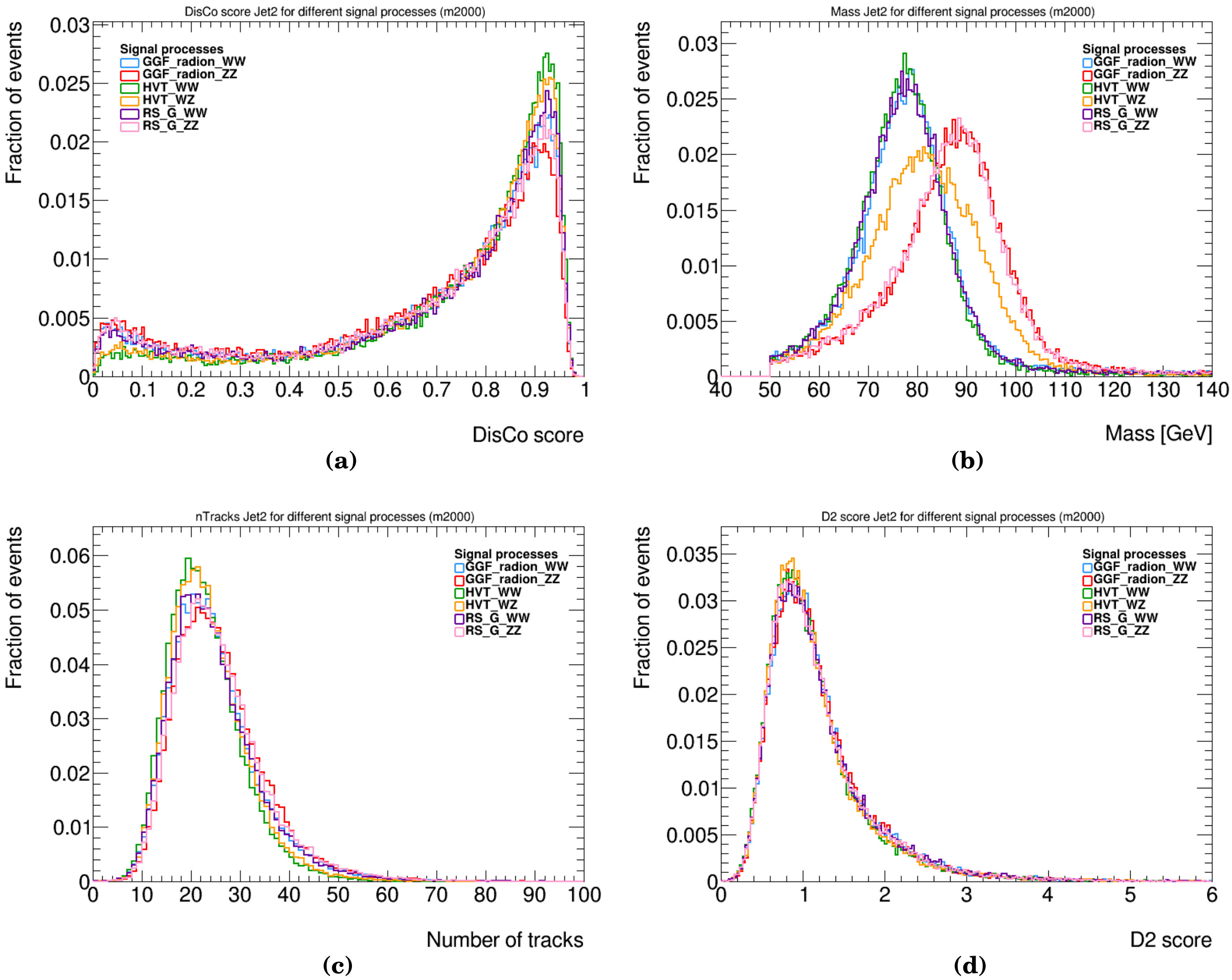


Figure 38: Distributions of the most discriminative secondary jet features for various signal models, including the Radion Model, Heavy Vector Triplet Model, and Bulk Randall-Sundrum Model. Most features demonstrate consistency across these models, with minor variations in jet mass reflecting the distinct vector boson configurations (WW, WZ, ZZ) associated with each signal.

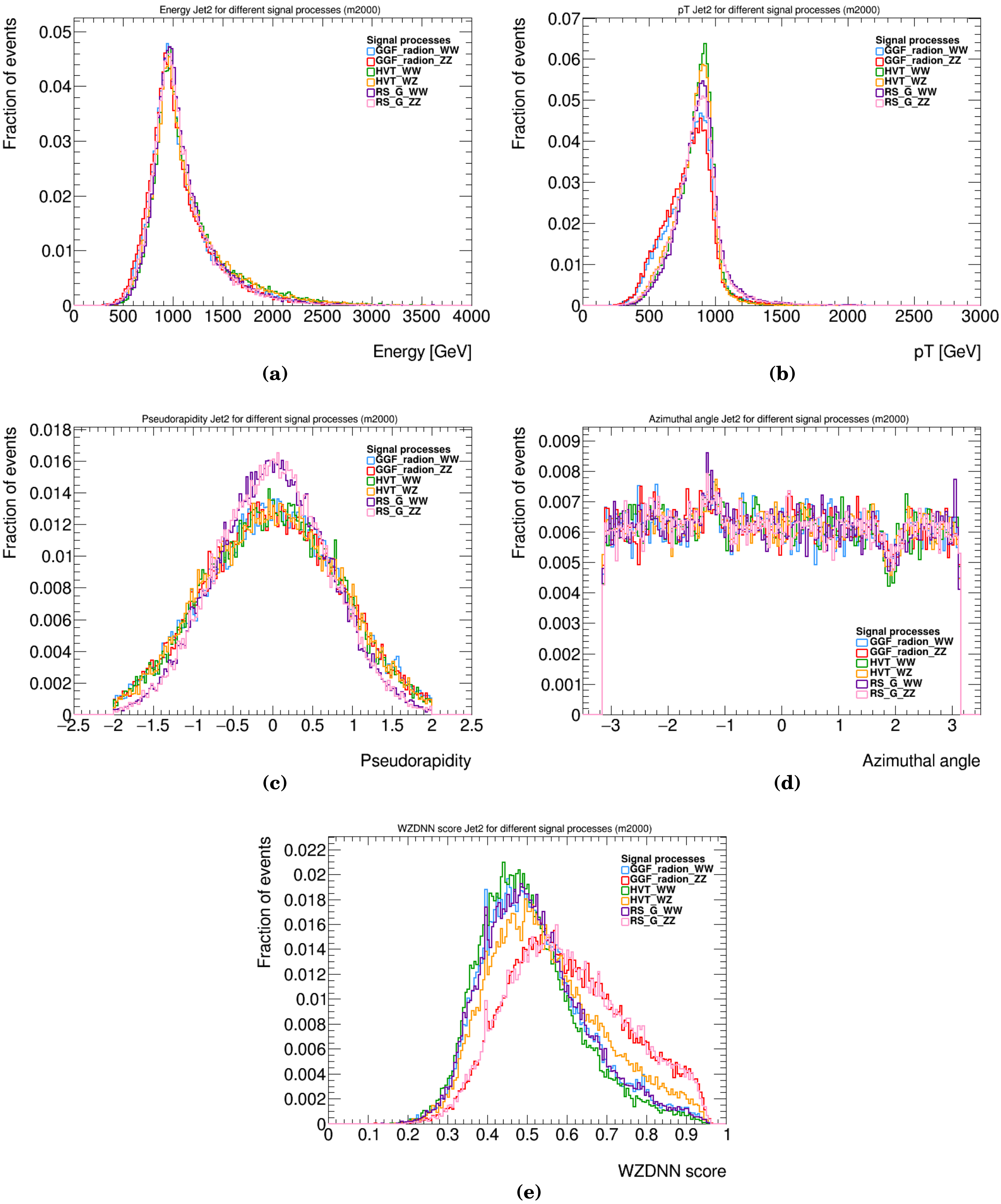


Figure 39: Distributions of additional secondary jet features for various signal models, including the Radion Model, Bulk Randall-Sundrum Model, and Heavy Vector Triplet Model. Most features exhibit consistency across these models, with minor variations in the WZDNN score reflecting the distinct vector boson configurations (WW, WZ, ZZ) associated with each signal.

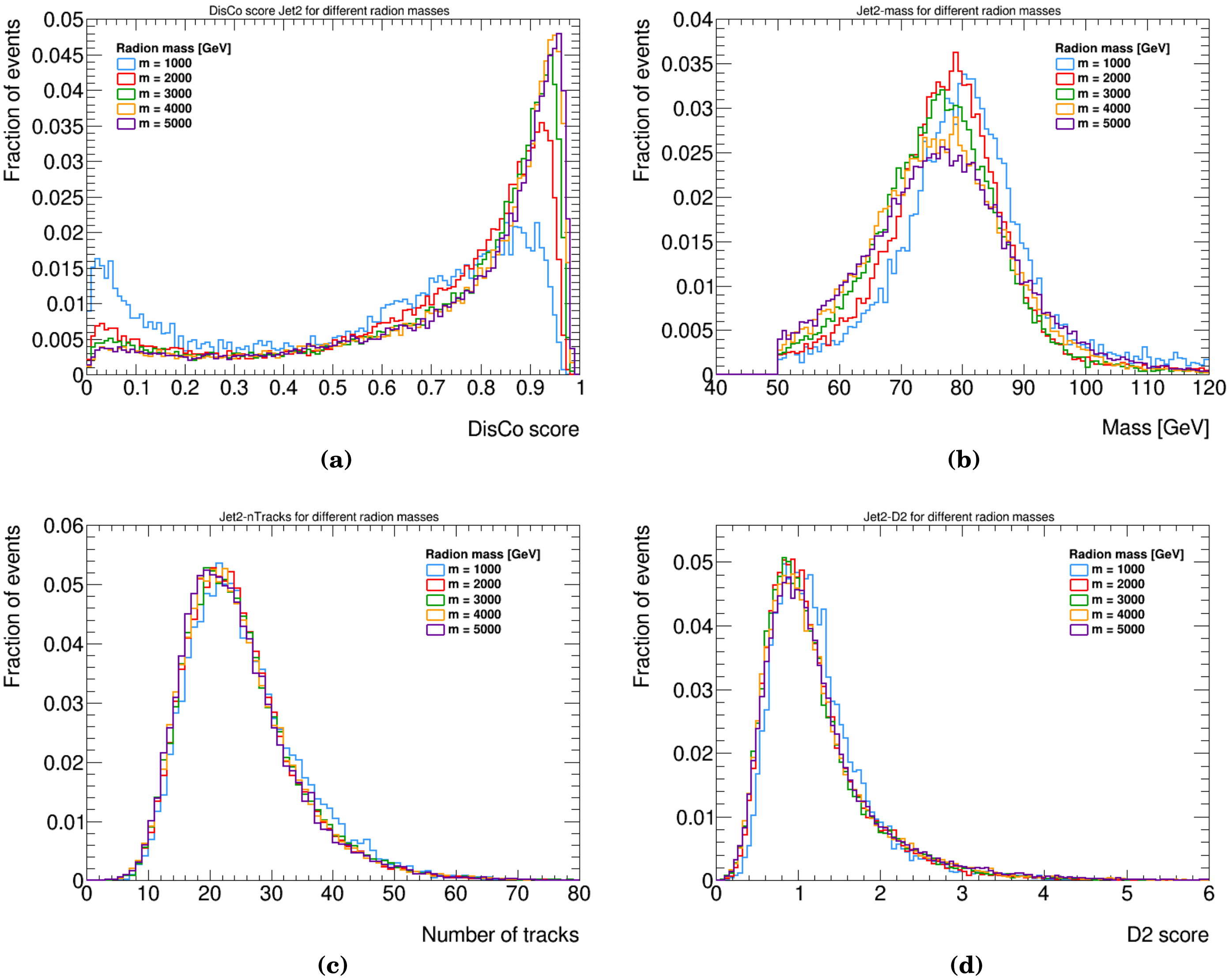
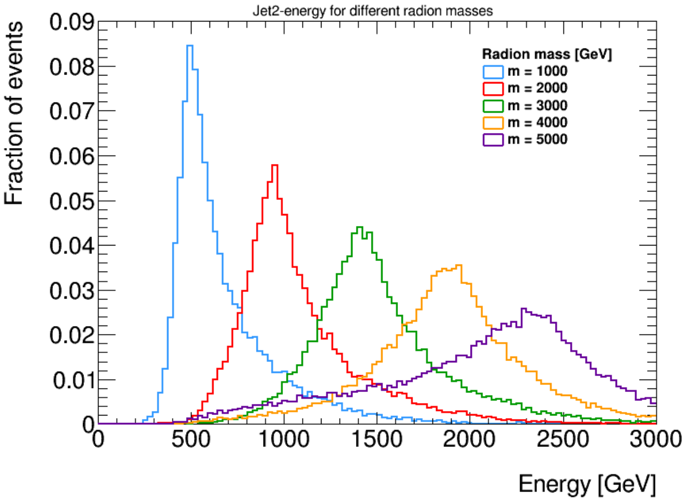
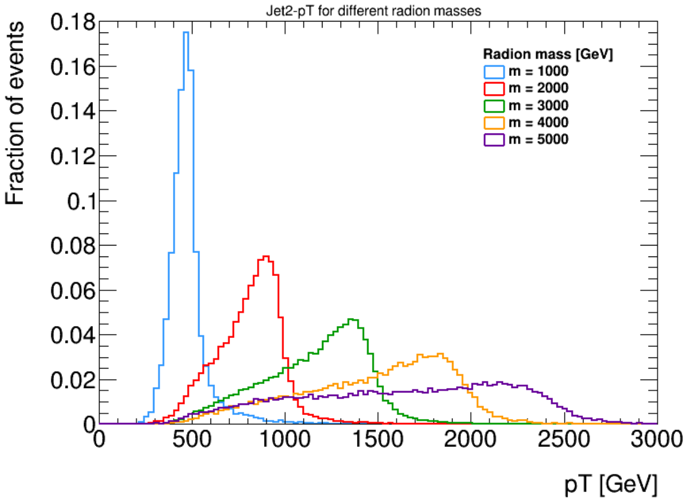


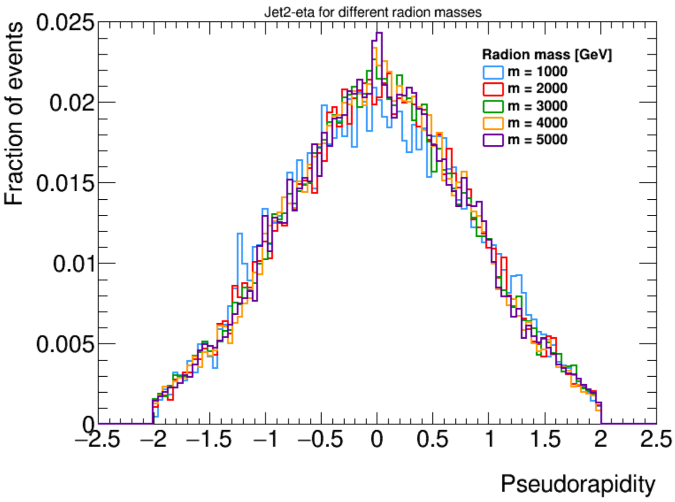
Figure 40: Distributions of the most discriminative secondary jet features for the Radion model signal at various resonance masses. The features, which include the DisCo score, jet mass, number of tracks, and D2 score, demonstrate consistency across different resonance masses.



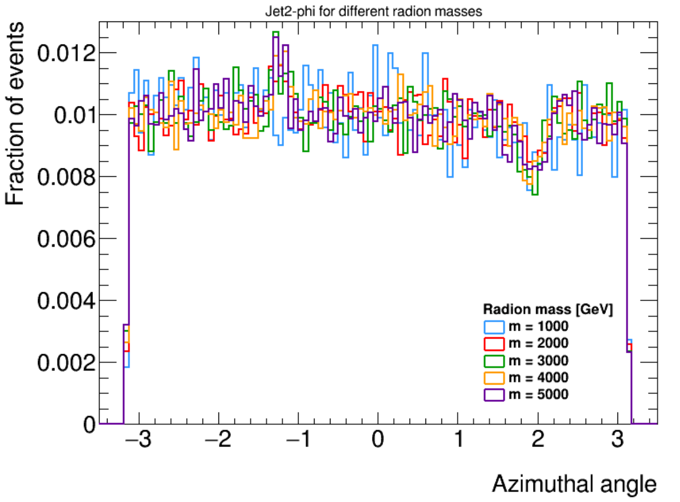
(a)



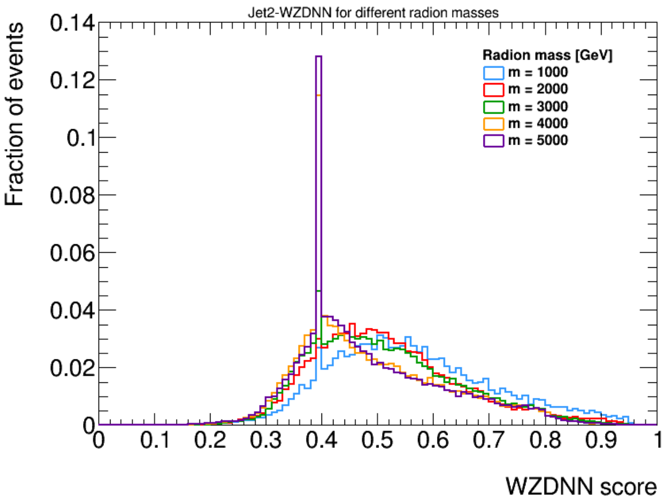
(b)



(c)



(d)



(e)

Figure 41: Distributions of additional secondary jet features for the Radion model signal at various resonance masses. Panels (a) and (b) illustrate how the energy and transverse momentum (p_T) scale with resonance mass, indicating their connection to the vector bosons decaying from the heavy resonance. The other features remain consistent across different resonance masses.

Following this, the discriminatory power of dijet features is explored. Figure 42 presents the most discriminative dijet features for the Radion model (m = 2000 GeV) against background processes. Features such as dijet energy and mass reveal substantial separation between signal and background, making them valuable for classification purposes. Figure 43 expands this analysis to include additional angular dijet features.

Figures 44 and 45 compare dijet features across the various signal models seen before. Notable deviations are observed for the HVT model, particularly in dijet p_T, demonstrating how various signal models influence feature distributions. Figures 46 and 47 further examine dijet features, focusing on the scaling of energy and mass with resonance mass. The m = 1000 GeV Radion exhibits unique behaviour in the pseudorapidity and rapidity distributions due to the imposed p_T cuts, which limit variation in these features.

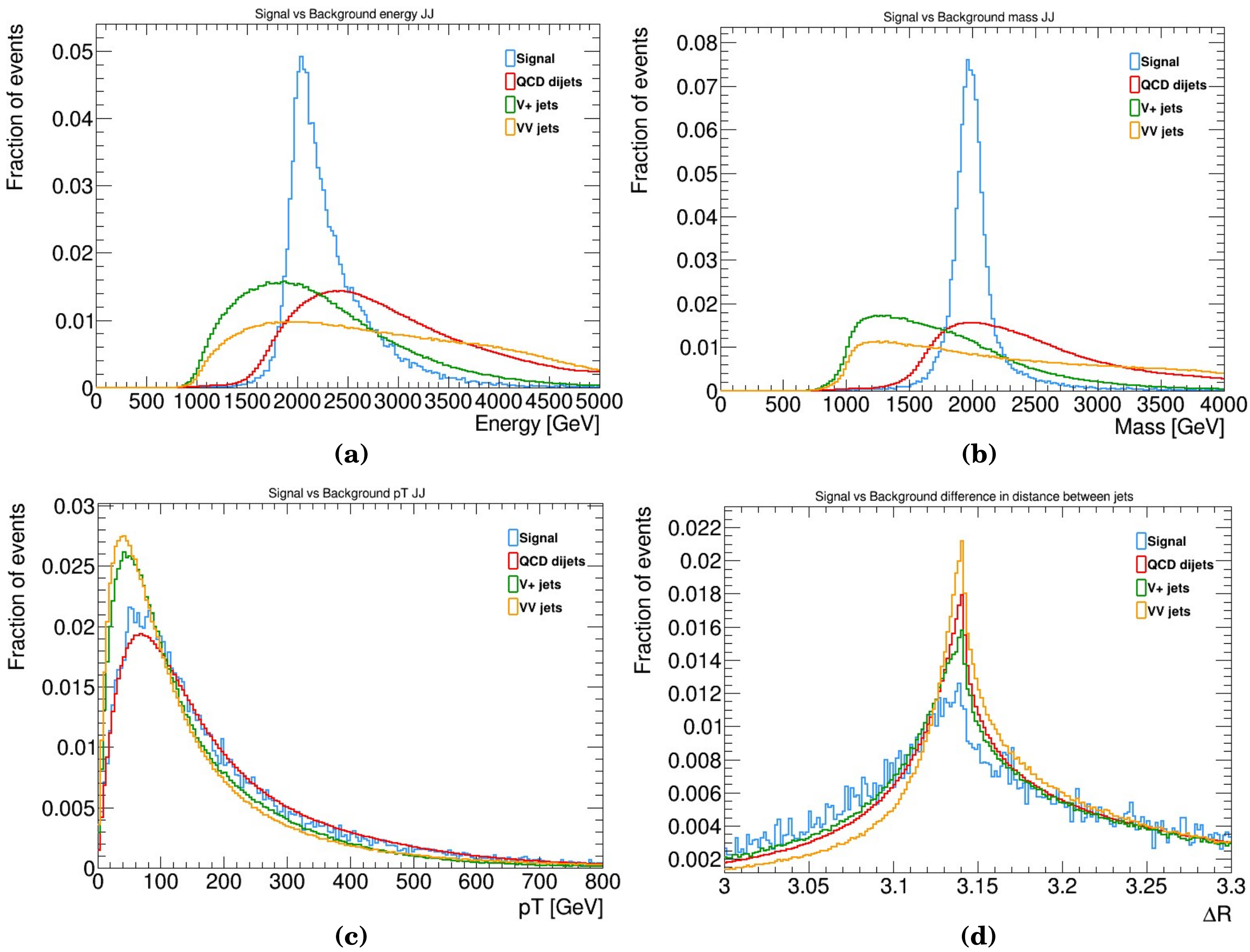


Figure 42: Distributions of some of the most discriminative dijet features for the Radion model signal at m = 2000 GeV, compared to various background processes. The features include dijet energy, mass, transverse momentum (p_T) and the distance between jets (ΔR).

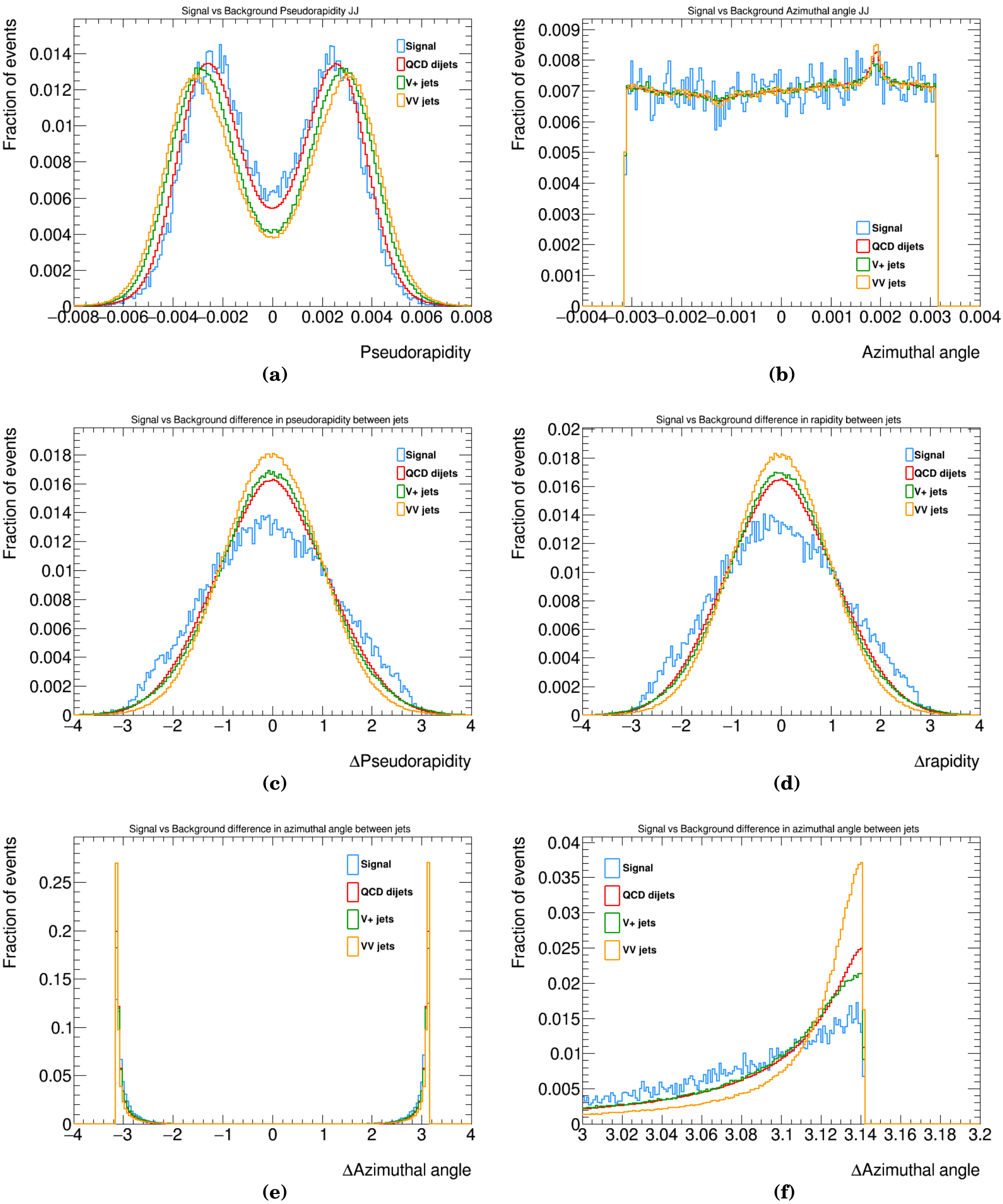


Figure 43: Distributions of additional dijet features for the Radion model signal at m = 2000 GeV, compared to various background processes. The features include pseudorapidity, azimuthal angle, and the differences in pseudorapidity ($\Delta\eta$), rapidity (Δy) and azimuthal angle ($\Delta\phi$).

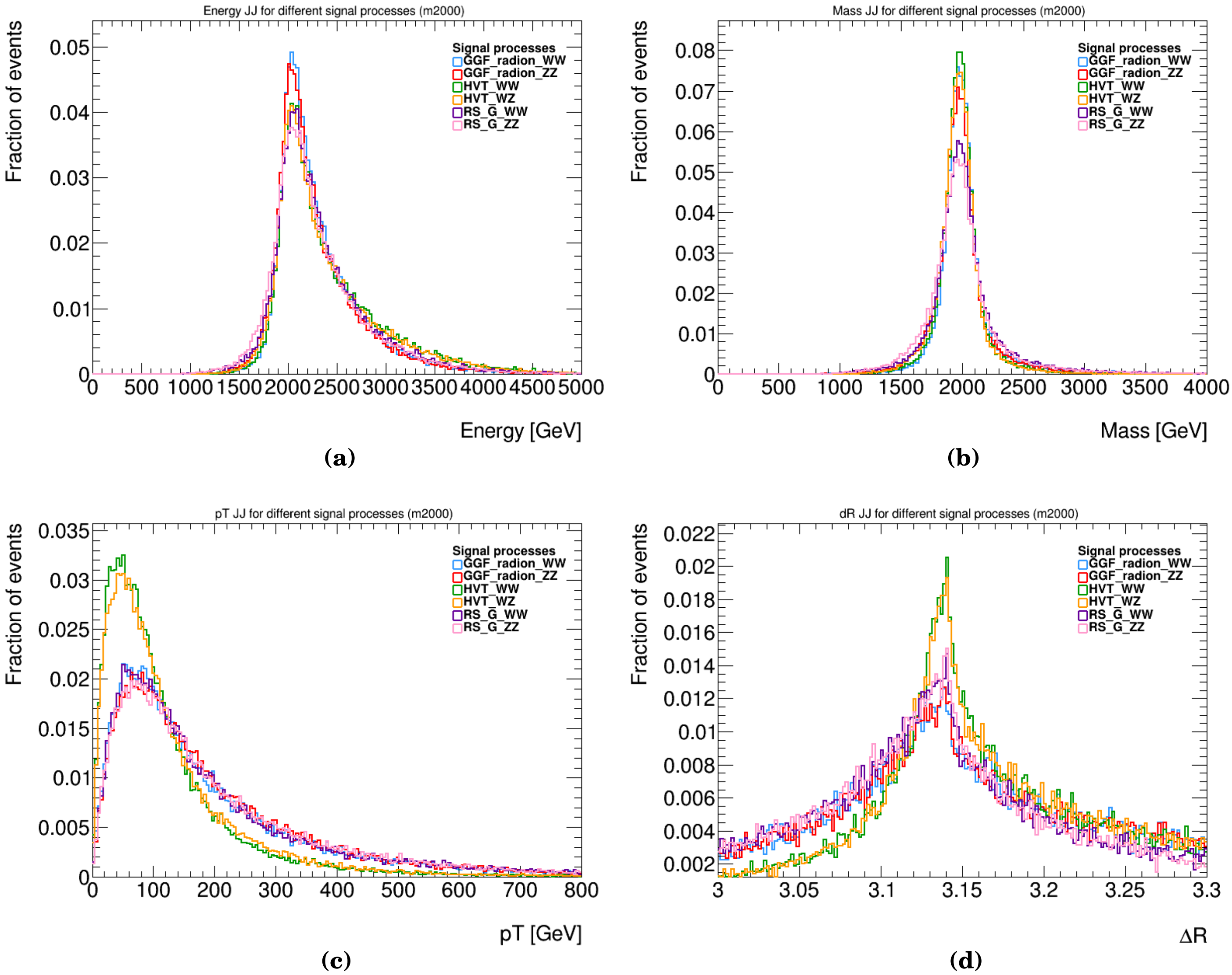


Figure 44: Distributions of the most discriminative dijet features for various signal models, including the Radion Model, Heavy Vector Triplet (HVT) Model, and Bulk Randall-Sundrum Model. The plots illustrate minor variations in features across the models, with the HVT model exhibiting notable deviations compared to the others, especially in the dijet transverse momentum (p_T).

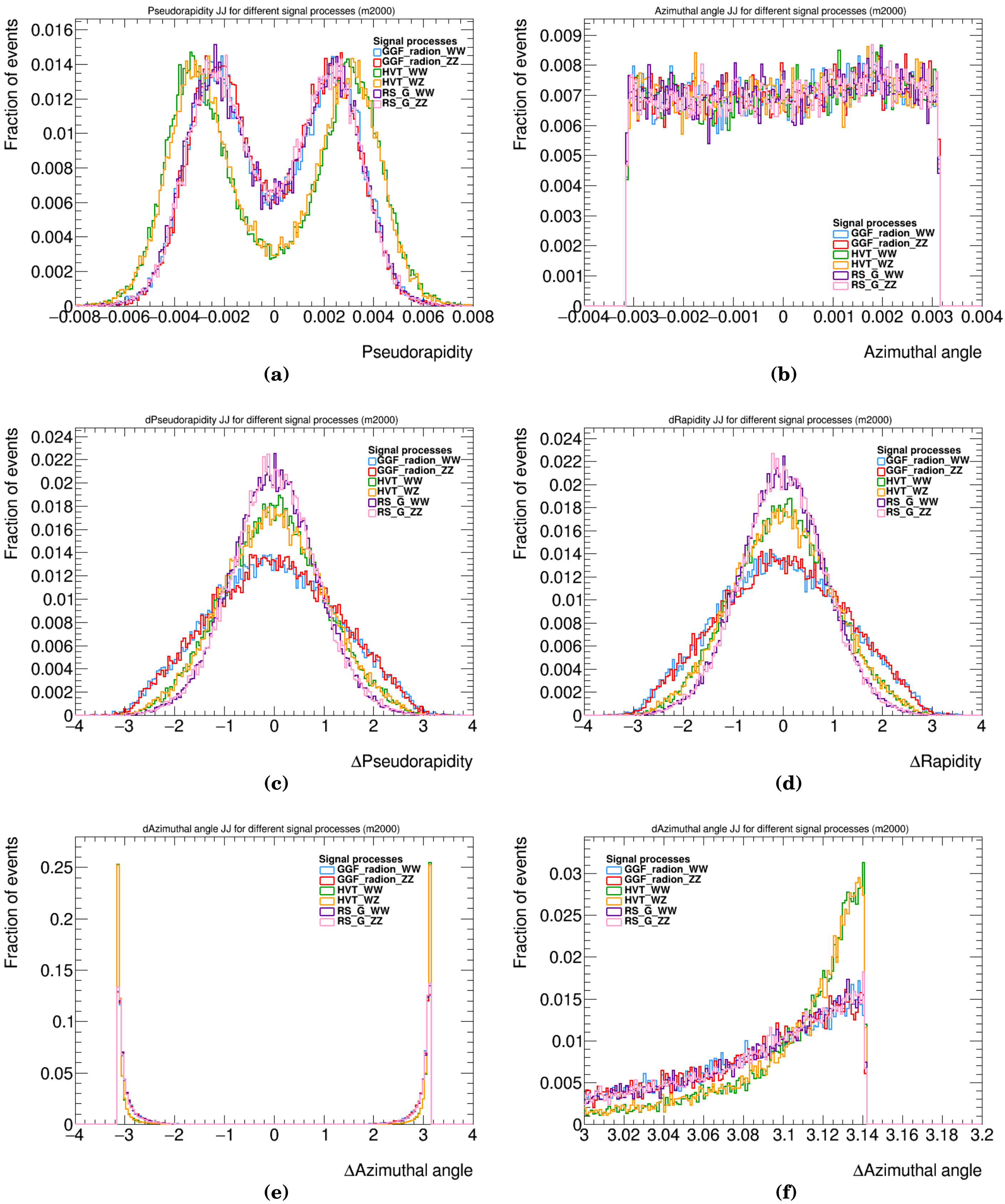


Figure 45: Distributions of additional dijet features for various signal models, including the Radion Model, Heavy Vector Triplet (HVT) Model, and Bulk Randall-Sundrum Model. The plots reveal minor variations in features among the models, with the HVT model showing significant deviations from the others. Additionally, the distributions highlight separations in pseudorapidity ($\Delta\eta$) and rapidity (Δy) across all models.

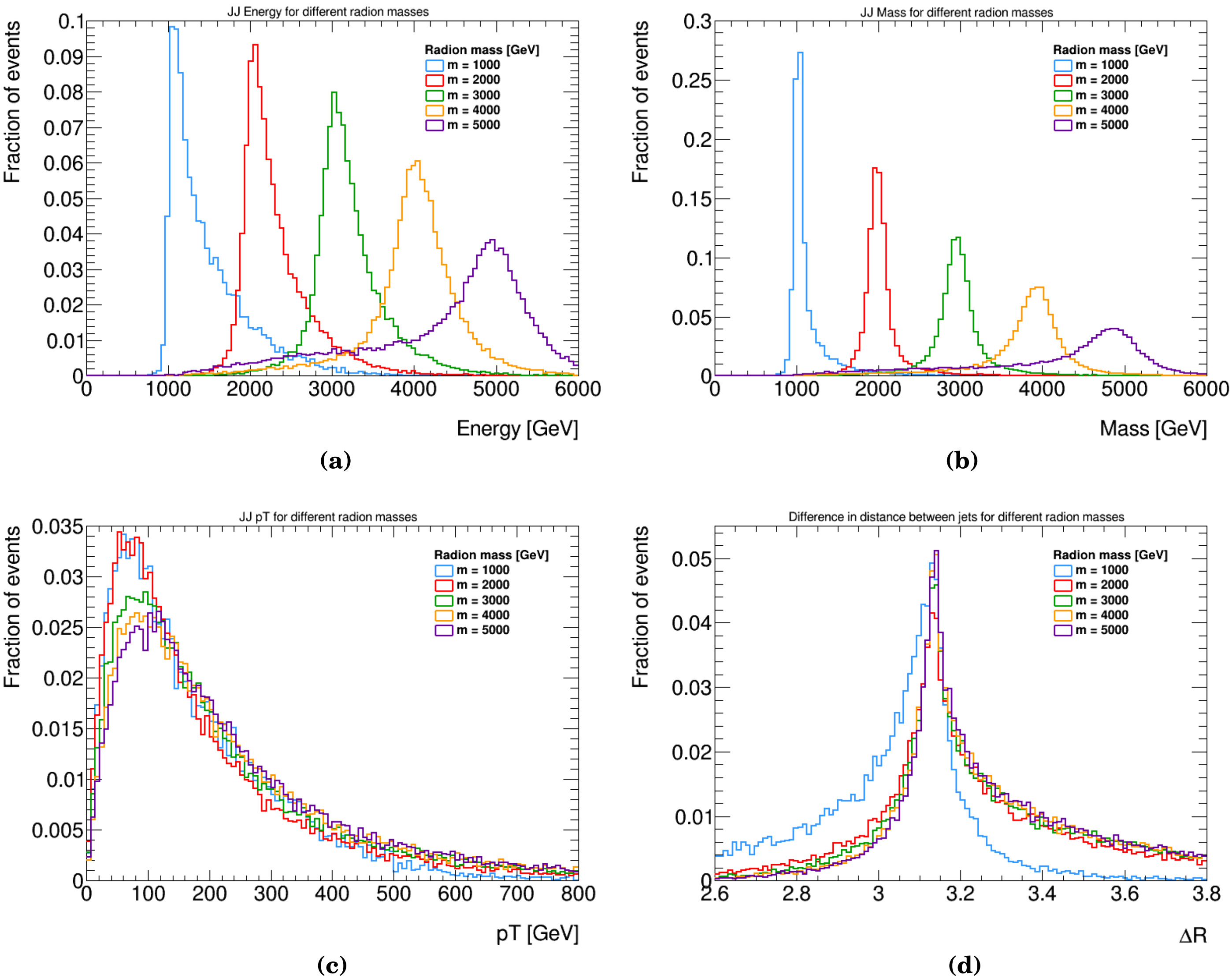


Figure 46: Distributions of some of the most discriminative dijet features for the Radion model signal at various resonance masses. Panels (a) and (b) illustrate how the dijet features of energy and mass scale with the heavy resonance mass. In contrast, panels (c) and (d) demonstrate that the transverse momentum (p_T) and the distance between jets (ΔR) remain largely unaffected and consistent across the different mass ranges.

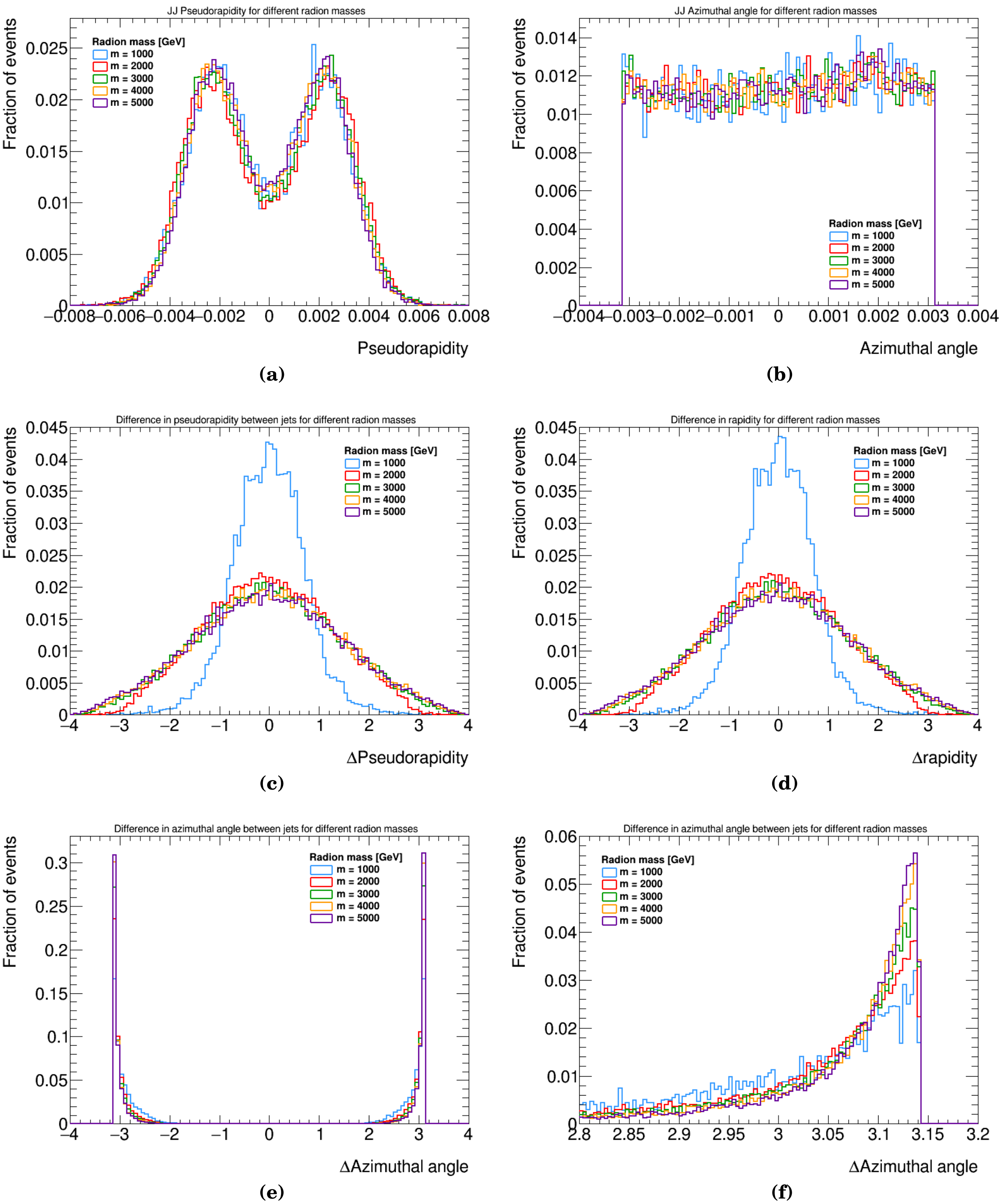
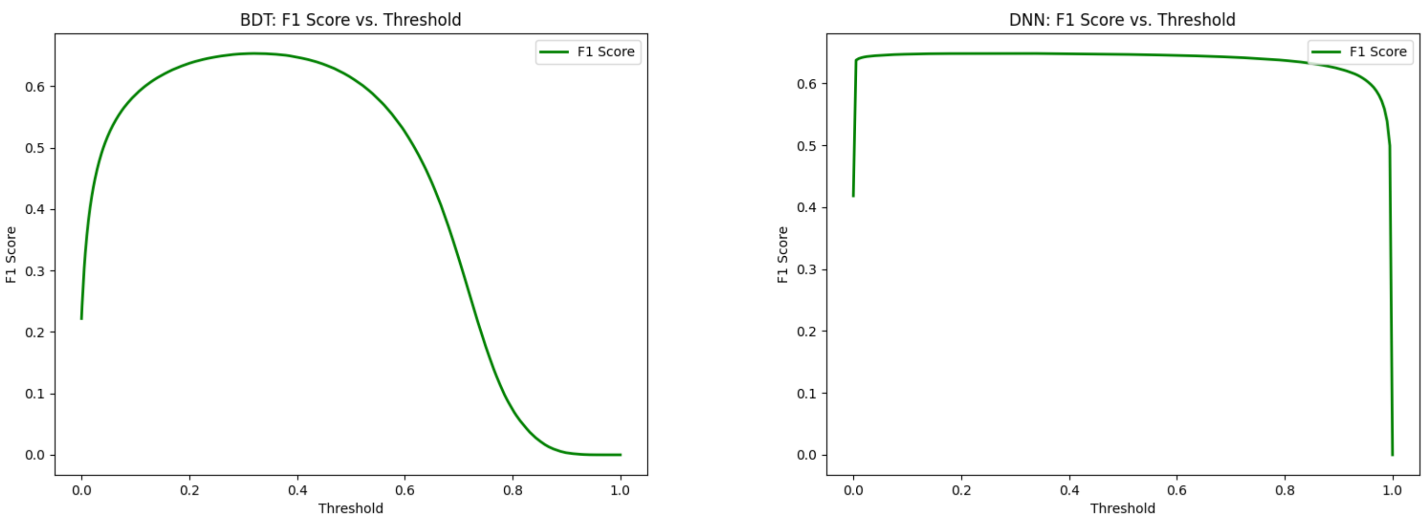


Figure 47: Distributions of additional dijet features for the Radion model signal at various resonance masses. The panels show that dijet features are generally consistent across different mass ranges, except for panels (c) and (d), which display a discrepancy for the m = 1000 GeV Radion. This can be explained by the jet p_T cut at 500 GeV, which restricts the data to the higher p_T half for this mass, resulting in less variation in pseudorapidity and rapidity, as seen in the plots.

The final part of this appendix examines the classification performance metrics for both the boosted decision tree (BDT) and deep neural network (DNN) models. Figure 48 compares the F1 score as a function of the classification threshold, reflecting their ability to balance precision and recall. This plot highlights the superior consistency of the DNN across a broader threshold range.

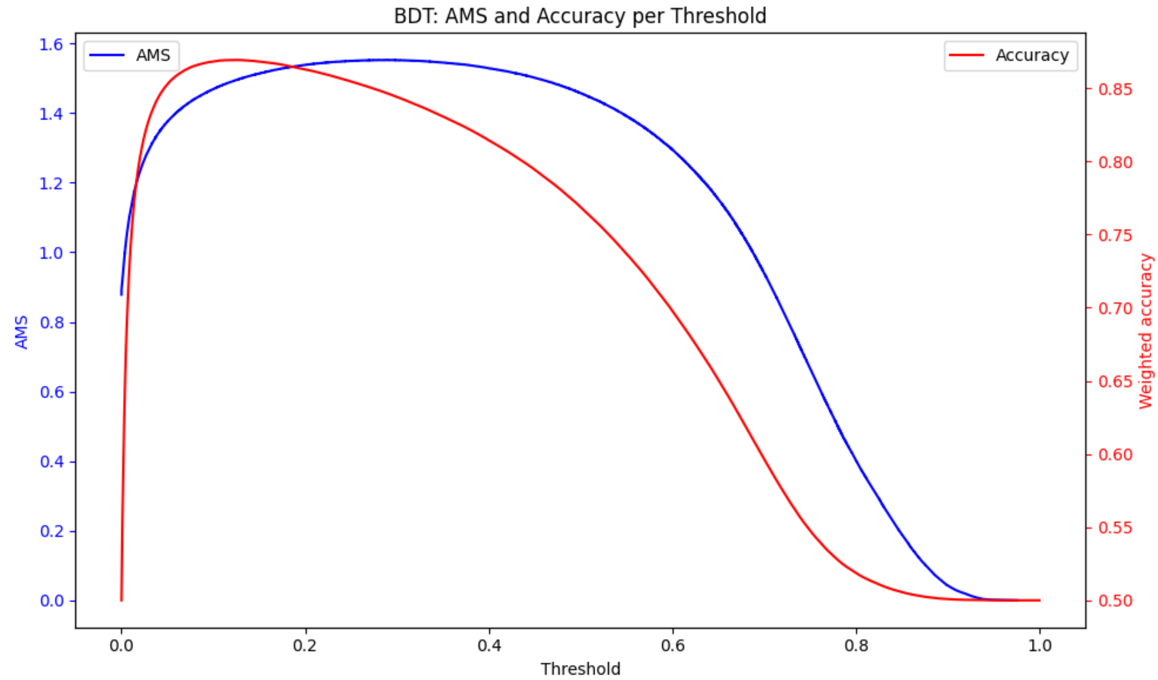
Additionally, Figure 49 compares the AMS (Approximate Median Significance) and weighted accuracy. Both models exhibit a peak in AMS around a threshold of 0.2, but the DNN again shows higher accuracy across a wider range of thresholds. Finally, Figure 50 presents the trade-offs in signal and background classification accuracy as the threshold is adjusted. While the BDT shows a rapid decline in signal accuracy, the DNN demonstrates a more gradual decline, with background accuracy remaining consistently high, indicating greater stability in classification performance.

In conclusion, the plots in this appendix provide additional context and clarity regarding the feature distributions and validation metrics for the signal vs. background classification.

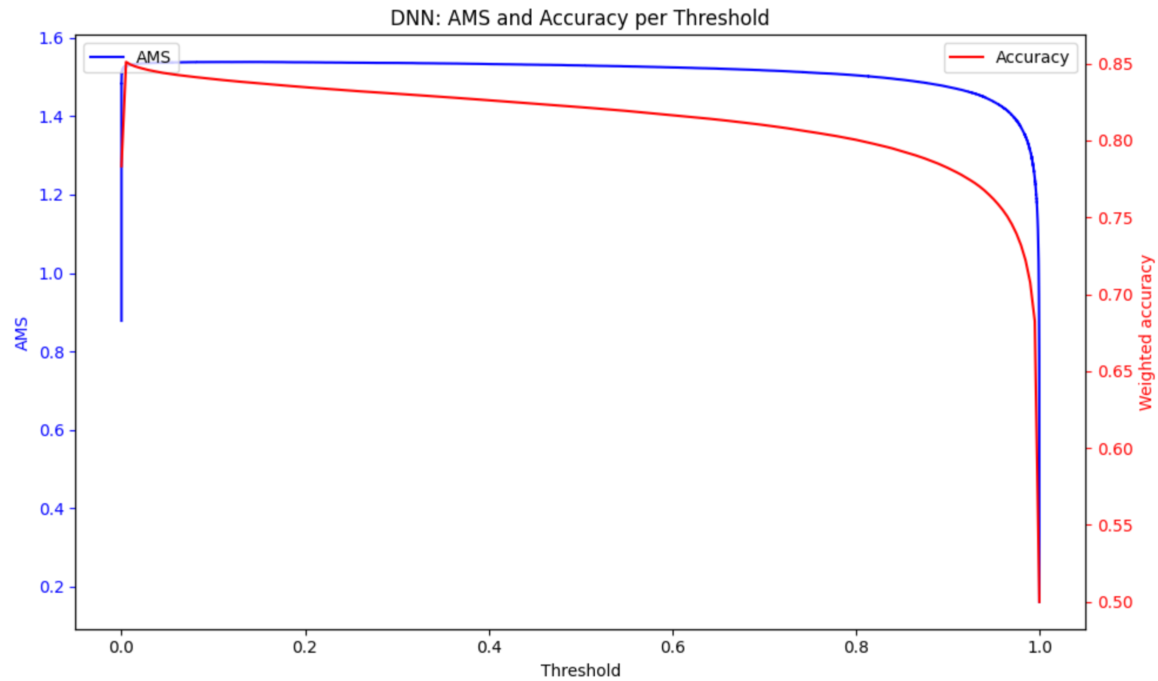


(a) **(b)**

Figure 48: Comparison of F1 score versus threshold for the boosted decision tree (BDT) and deep neural network (DNN) models. The F1 score, a harmonic mean of precision and recall, measures the balance between correctly identifying signal and distinguishing it from background. In plot (a), the BDT starts with a low F1 score around 0.2 at threshold zero, rising to a peak of approximately 0.65 between thresholds 0.3 and 0.4, then gradually declining, approaching zero after threshold 0.8. Conversely, in panel (b), the DNN begins at an F1 score of 0.4, rises sharply to 0.65, and remains flat up to threshold 0.8, after which it gradually decreases and falls sharply near threshold 1.0. This indicates that the DNN maintains more consistent performance over a wider range of thresholds compared to the BDT.

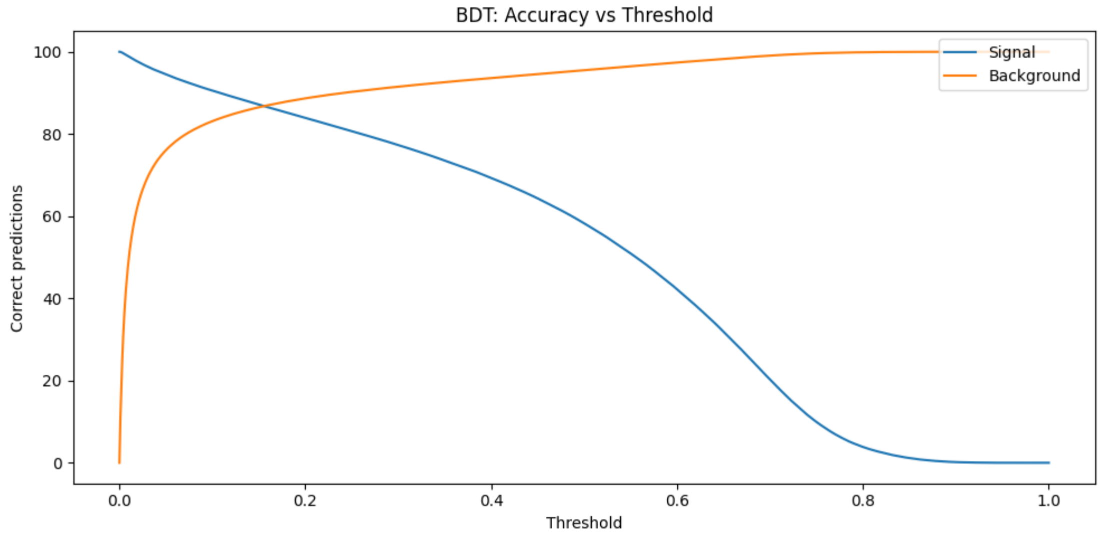


(a)

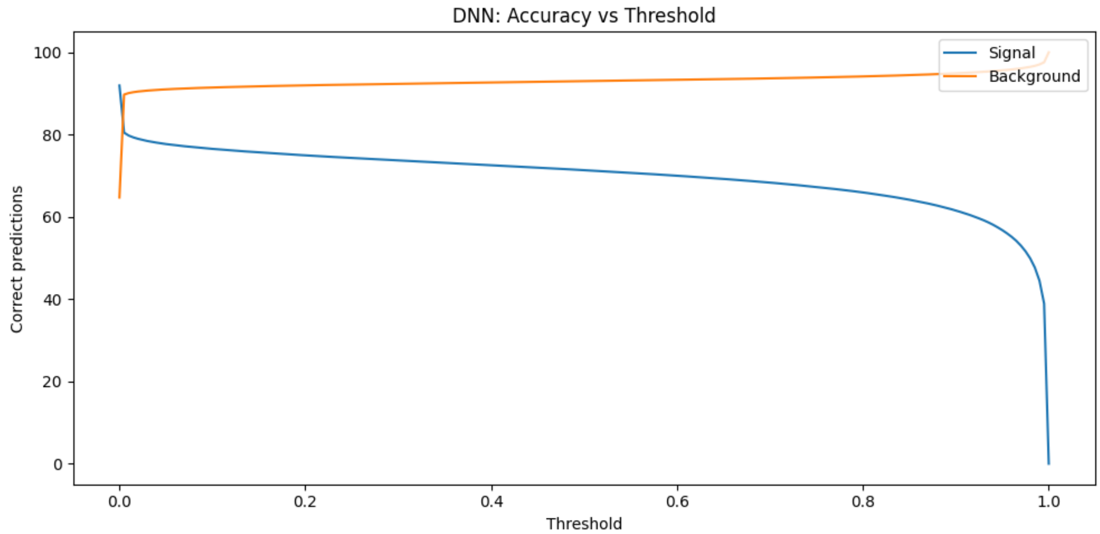


(b)

Figure 49: Comparison of performance metrics for the boosted decision tree (BDT) and deep neural network (DNN) models. The plots show the AMS (blue curve) and weighted accuracy (red curve) as a function of the classification threshold. AMS (Approximate Median Significance) is a metric used to evaluate the statistical significance of a signal in the presence of background noise, commonly applied in high-energy physics analyses. For the BDT (a), the AMS reaches its maximum around a threshold of 0.2, while the weighted accuracy ($\sqrt{Acc_{\text{signal}}^2 + Acc_{\text{bkg}}^2}$) decreases more gradually as the threshold increases. For the DNN (b), the AMS also peaks around 0.2 but declines more slowly, with the weighted accuracy remaining consistently higher across most thresholds. These plots highlight the models' sensitivity to threshold selection, impacting both AMS and weighted accuracy.



(a)



(b)

Figure 50: Comparison of signal and background classification accuracy for the boosted decision tree (BDT) and deep neural network (DNN) models as a function of the classification threshold. The upper plot (a) shows the accuracy for signal (blue) and background (orange) for the BDT model, where signal accuracy starts high and decreases with increasing threshold, while background accuracy increases rapidly before levelling off. In the lower plot (b), the DNN shows a more stable trend, with signal accuracy decreasing more gradually compared to the BDT, and background accuracy remaining consistently high across most threshold values. These plots demonstrate the trade-off between correctly identifying signal and background events as the classification threshold is adjusted.

Appendix C: Polarisation Related Plots

This appendix includes supplementary plots that offer deeper insights into the polarisation classification task, as well as additional $\cos\theta$ regression plots. The figures highlight distributions of various jet, subjet, and particle flow (pflow) features, along with several self-engineered features. Additionally, it presents further validation metrics for the machine learning models.

The analysis begins by examining the discrimination power of jet features in distinguishing polarisation states. Figure 51 displays the jet substructure (JSS) variables and various jet features, including $\mathrm{Split}_{2,3}$, Q_w, p_T, mass, η and ϕ. These features provide additional capabilities for differentiating between the characteristics of the two polarisation states. Figure 52 presents additional N-subjettiness features, which demonstrate limited discrimination between the polarisation states. Finally, Figure 53 showcases the remaining jet features, which largely exhibit similar distributions for both polarisation states.

Figures 54, 55 and 56 extend this analysis by comparing the distributions of subjet features. In particular, the mass, $\Delta\eta$ and $\Delta\phi$ features in the first figure reveal some discrimination between the polarisation states, especially for the secondary subjet. In contrast, the other two figures exhibit minimal variation, with the distributions overlapping almost perfectly for both polarisations.

Following this, the discriminatory power of the particle flow features is evaluated, focusing on the five most energetic particles. Figure 57 presents the angular particle flow features for both longitudinal and transverse polarisations. These features, however, show little to no discriminatory power. A similar result is observed in Figure 58, where the z_0 values are plotted, revealing minimal differences between the polarisations.

In the subsequent plots, Figures 59 and 60, the three-dimensional impact parameters (IP3D) of the pflow tracks are displayed. While there is some observable discrimination for the significance, the variance remains relatively small, limiting their effectiveness for classification. The final particle flow features (Figs. 61, 62), which include additional parameters such as the p_T fraction and the number of degrees of freedom, similarly offer no significant advantage in distinguishing between polarisations.

The next section delves into additional engineered features. Figure 63 presents the differences between the jet and the primary subjet features, including ΔE, Δp_T, ΔM, $\Delta\eta$ and $\Delta\phi$. The subsequent figures show the differences between the jet and the secondary subjet(Fig. 64), and between the two subjets (Fig. 65). Unfortunately, these figures reveal only limited separation between the polarisation states. On the other hand, Figure 66 illustrates the angular distance between the jet and subjets, showing noticeable discrimination, particularly for the second subjet.

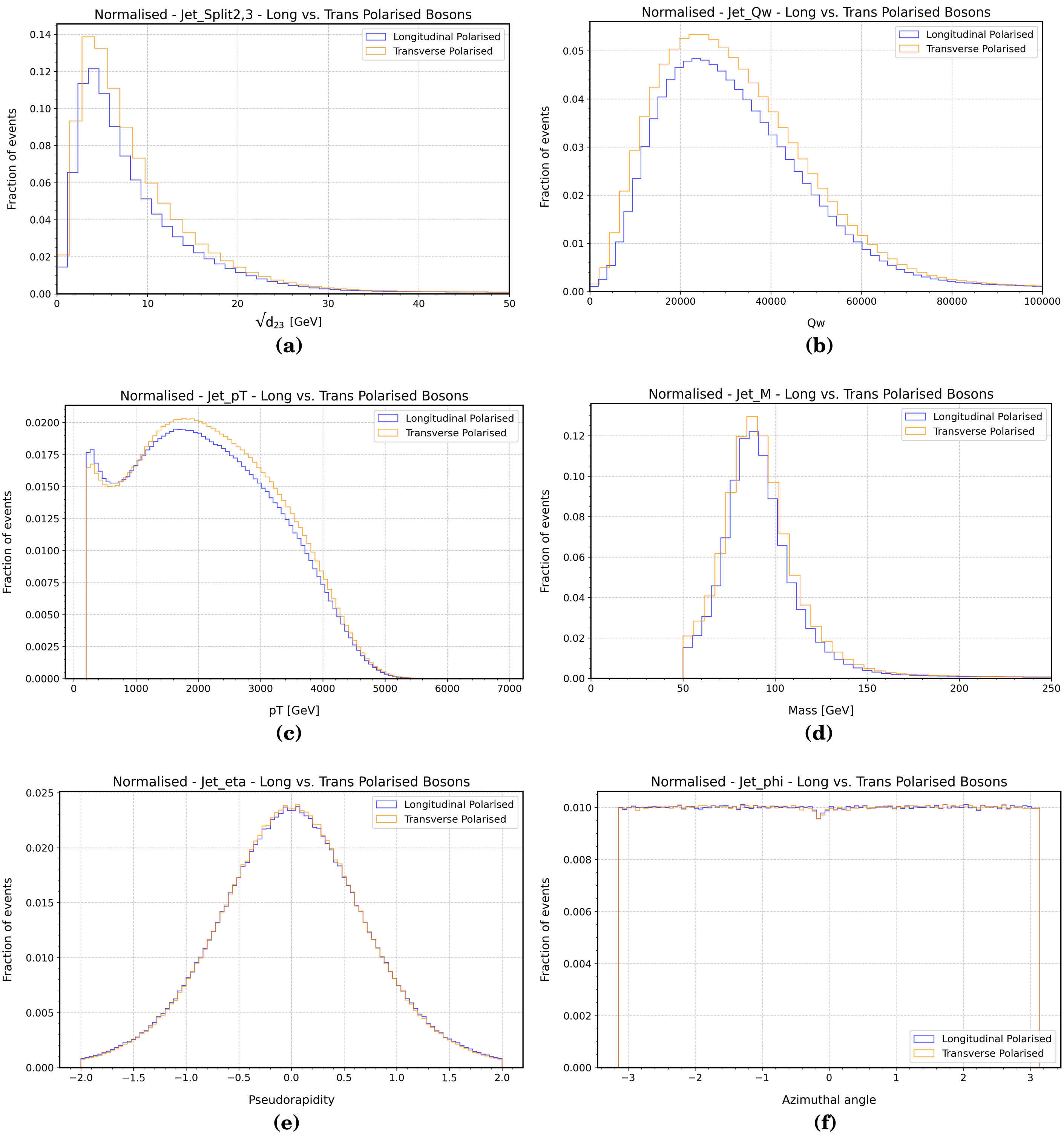


Figure 51: Distributions of additional jet features for polarisation classification, including the $\text{Split}_{2,3}$, Q_w, p_T, mass, eta and phi. Showing some separation but nothing substantial. These features demonstrate some separation between the polarisation states, though the differences are not substantial.

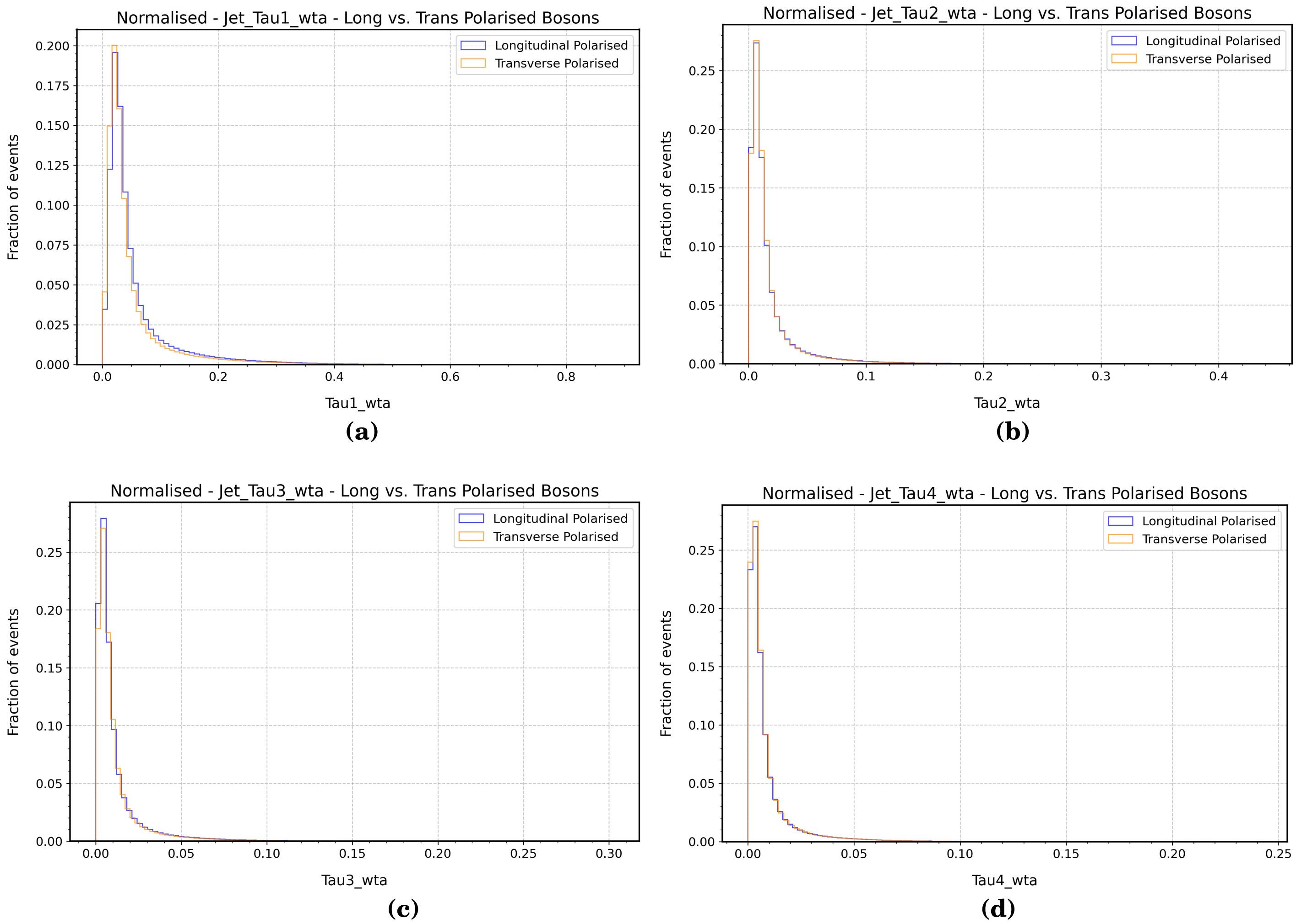


Figure 52: Distributions of N-subjettiness features for polarisation classification, including the τ_1_wta, τ_2_wta, τ_3_wta and τ_4_wta. These features seem to provide minimal separation between the polarisation states.

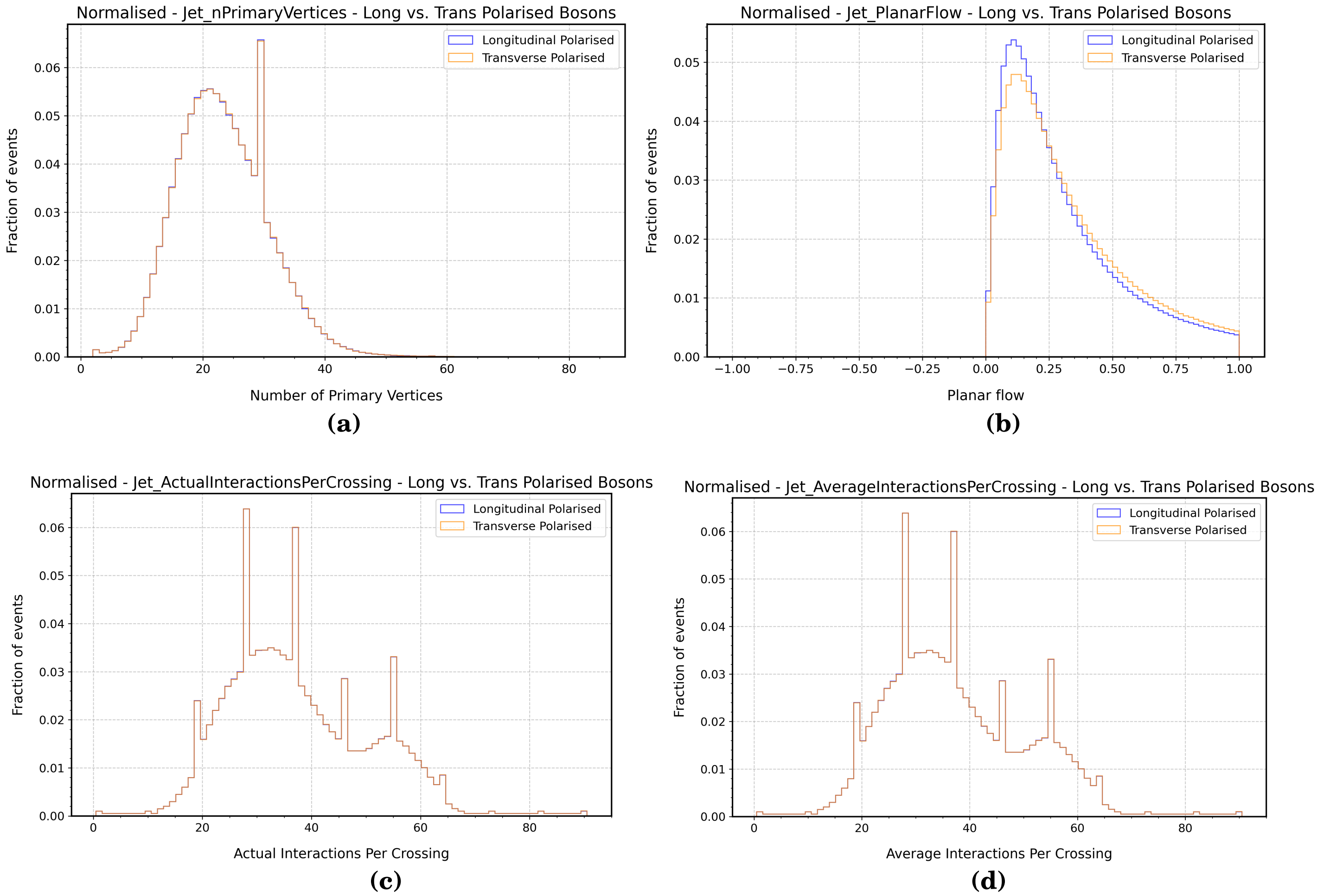


Figure 53: Plots of additional non-discriminative jet features for polarisation classification, including the number of primary vertices, planar flow, and actual and average interactions per crossing. The distributions for the different polarisation states largely overlap.

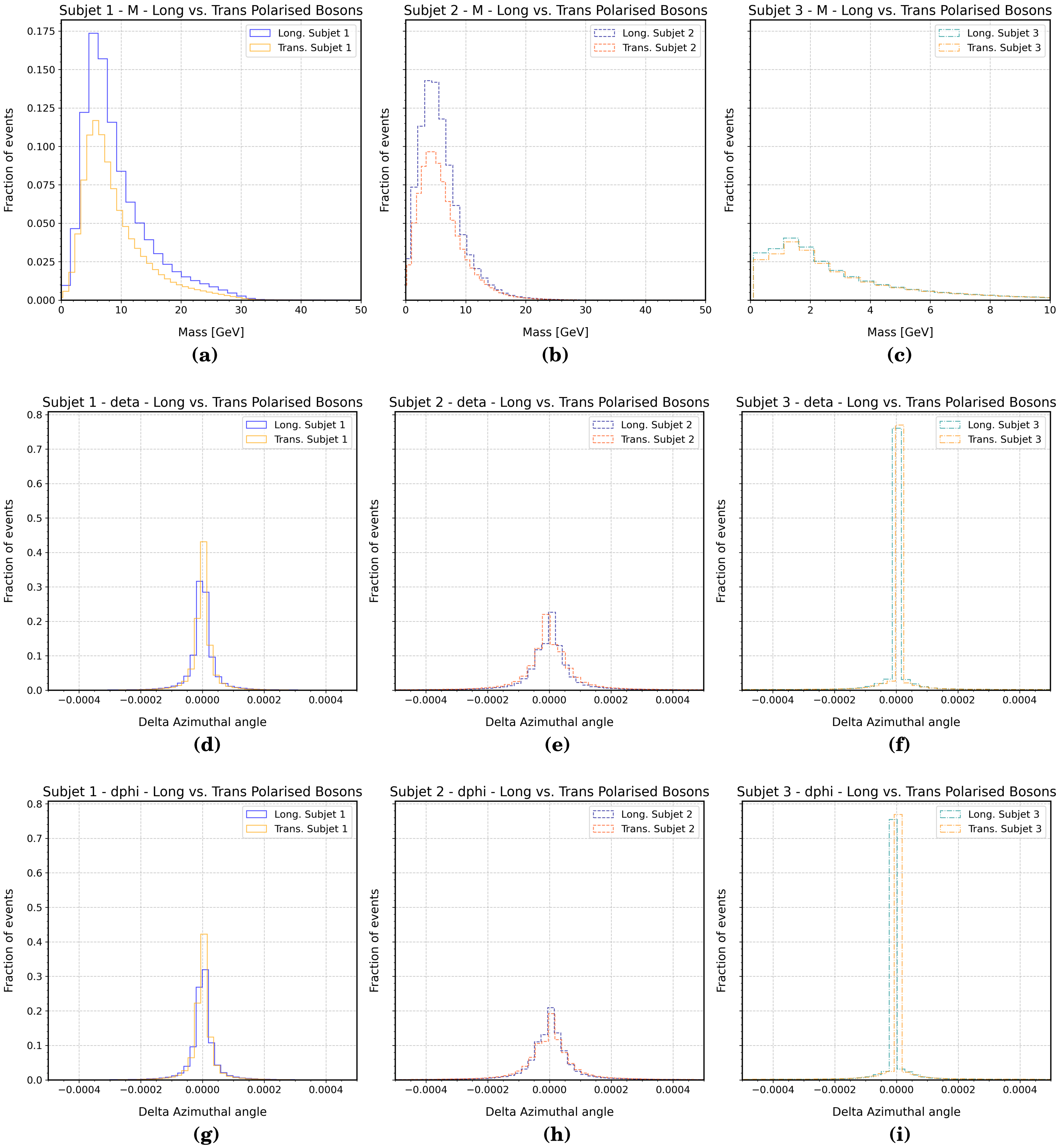


Figure 54: Distributions of additional discriminative subjet features for polarisation tagging, including mass, $\Delta\eta$ and $\Delta\phi$ for all subjets, with the secondary subjet providing the highest discrimination between polarisation states.

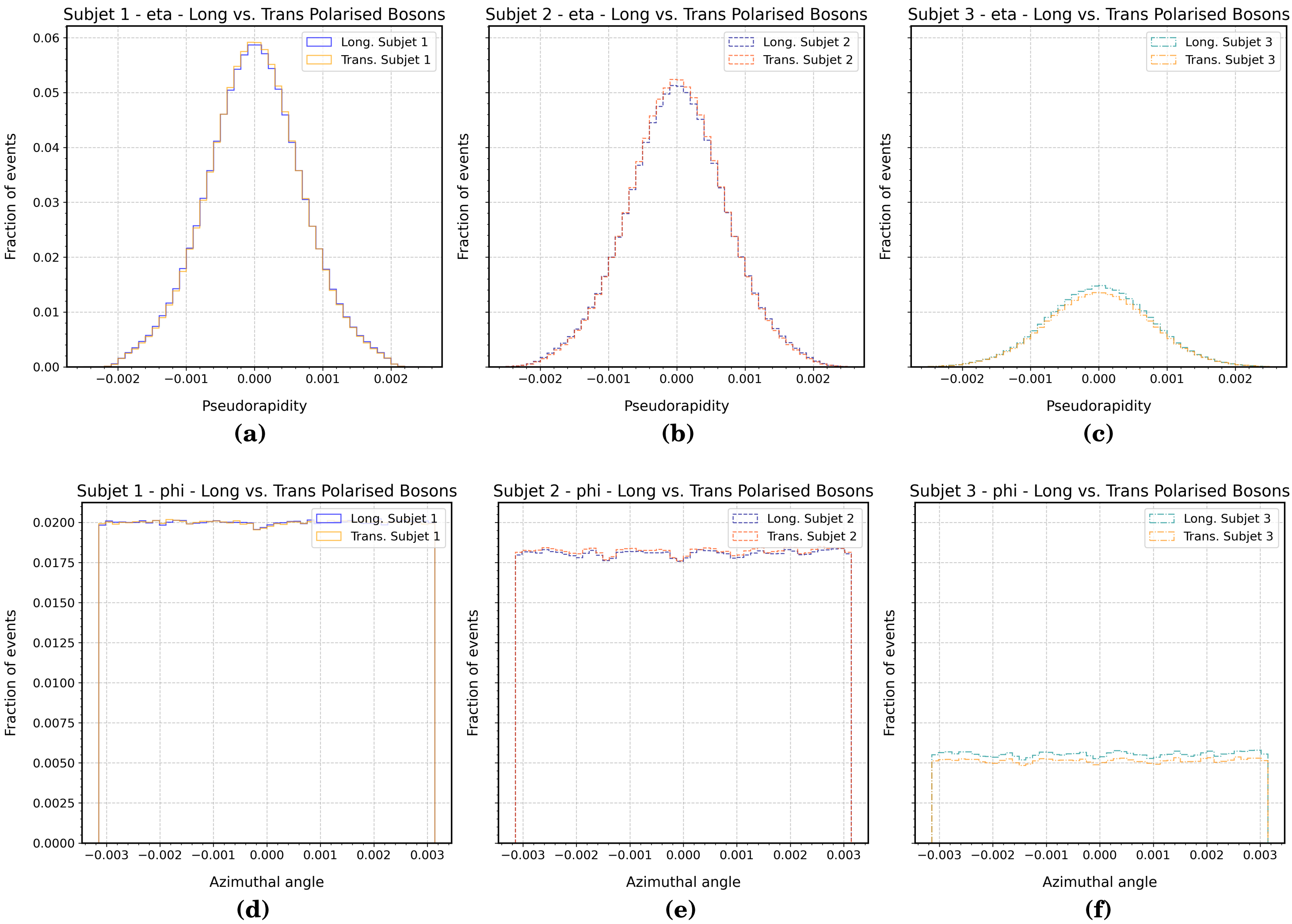


Figure 55: Distributions of additional angular subjet features for polarisation tagging, including pseudorapidity (η) and azimuthal angle (ϕ) for all subjets. Minimal discrimination is observed from these features.

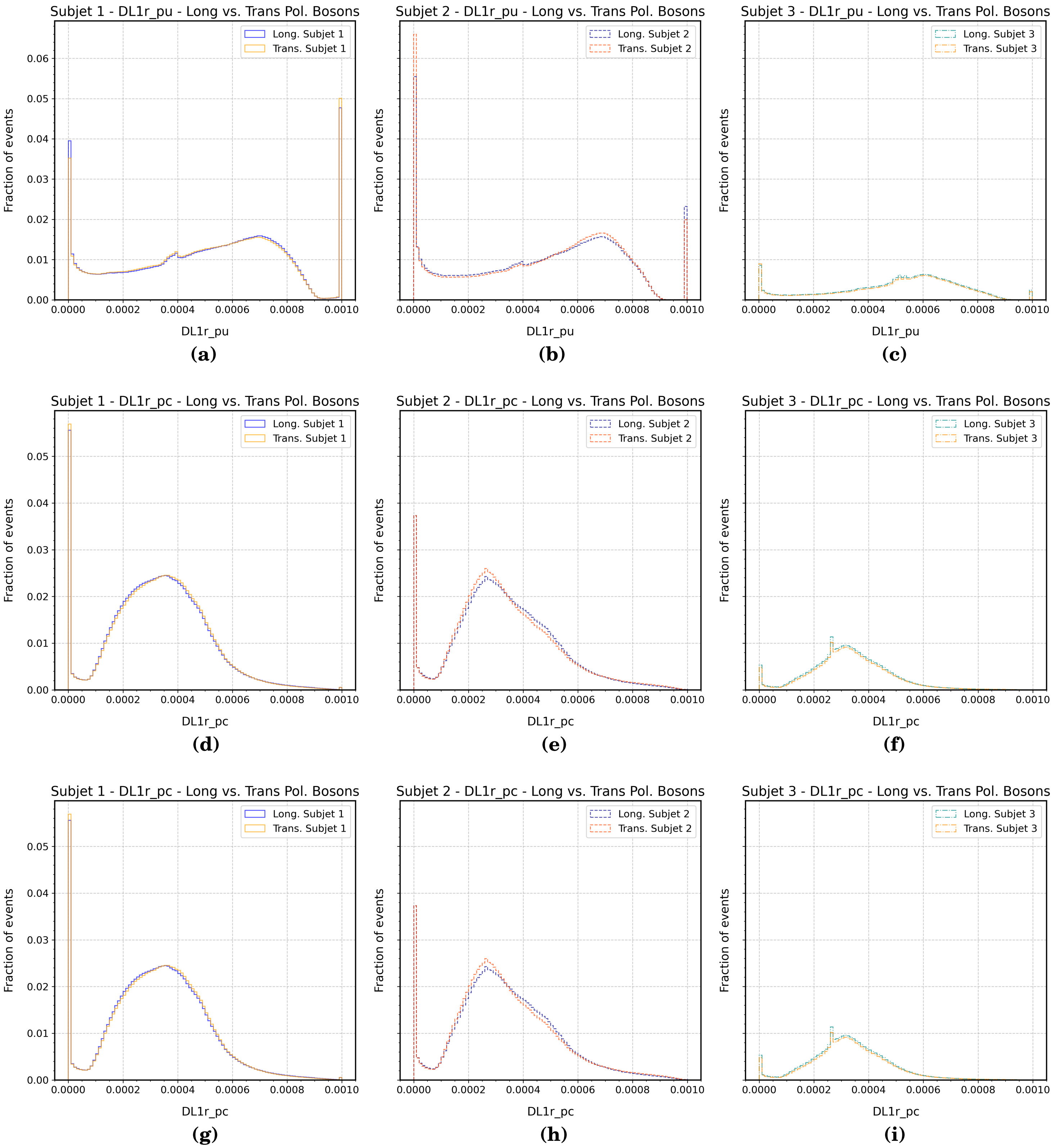


Figure 56: Distributions of additional non-discriminative subjet information for polarisation tagging based on deep neural network flavour tagging, where the distributions of both polarisation states largely overlap.

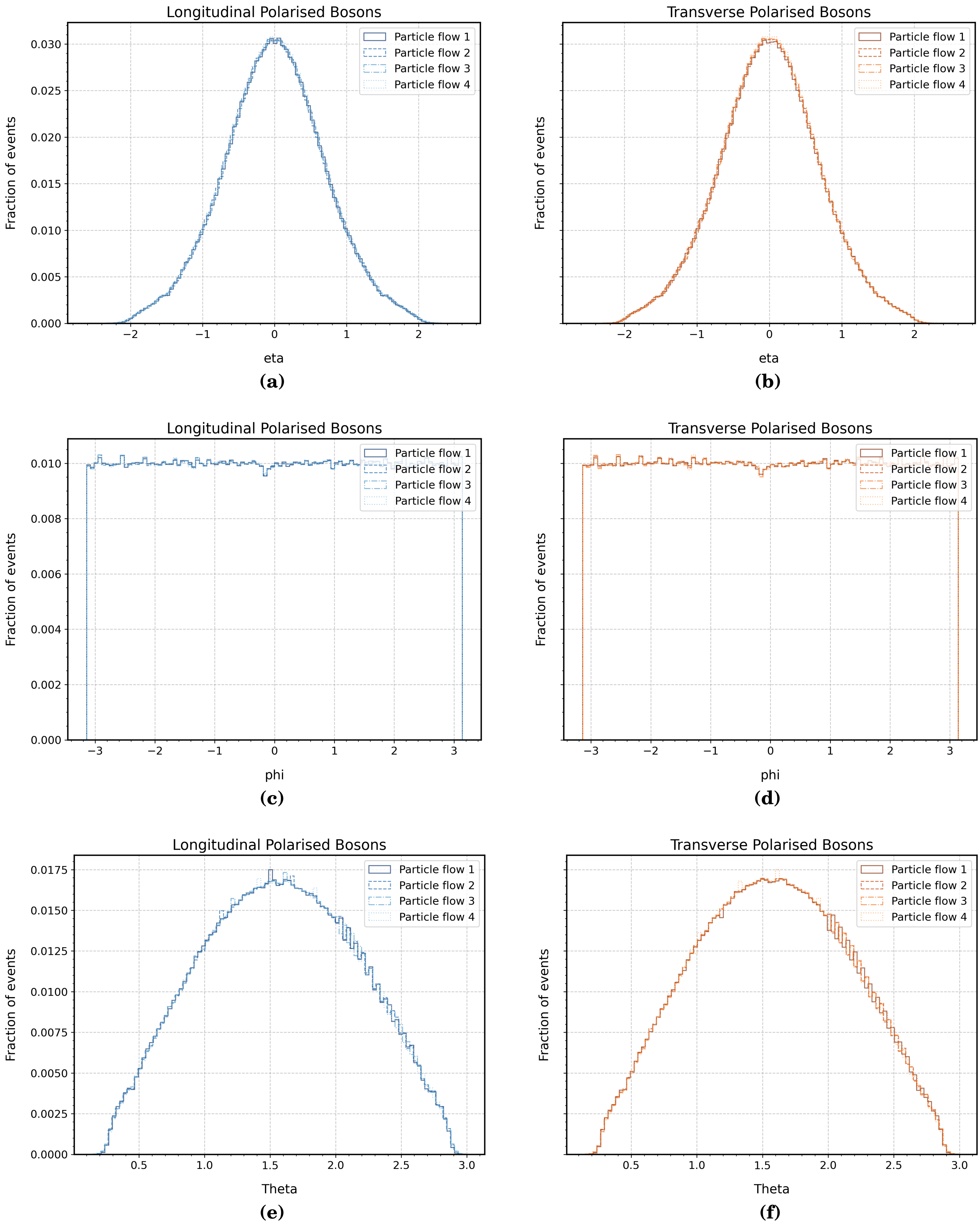


Figure 57: Distributions of non-discriminative angular particle flow features for polarisation tagging, including η, ϕ and θ. The plots on the left (longitudinal) and right (transverse) exhibit similar distribution shapes.

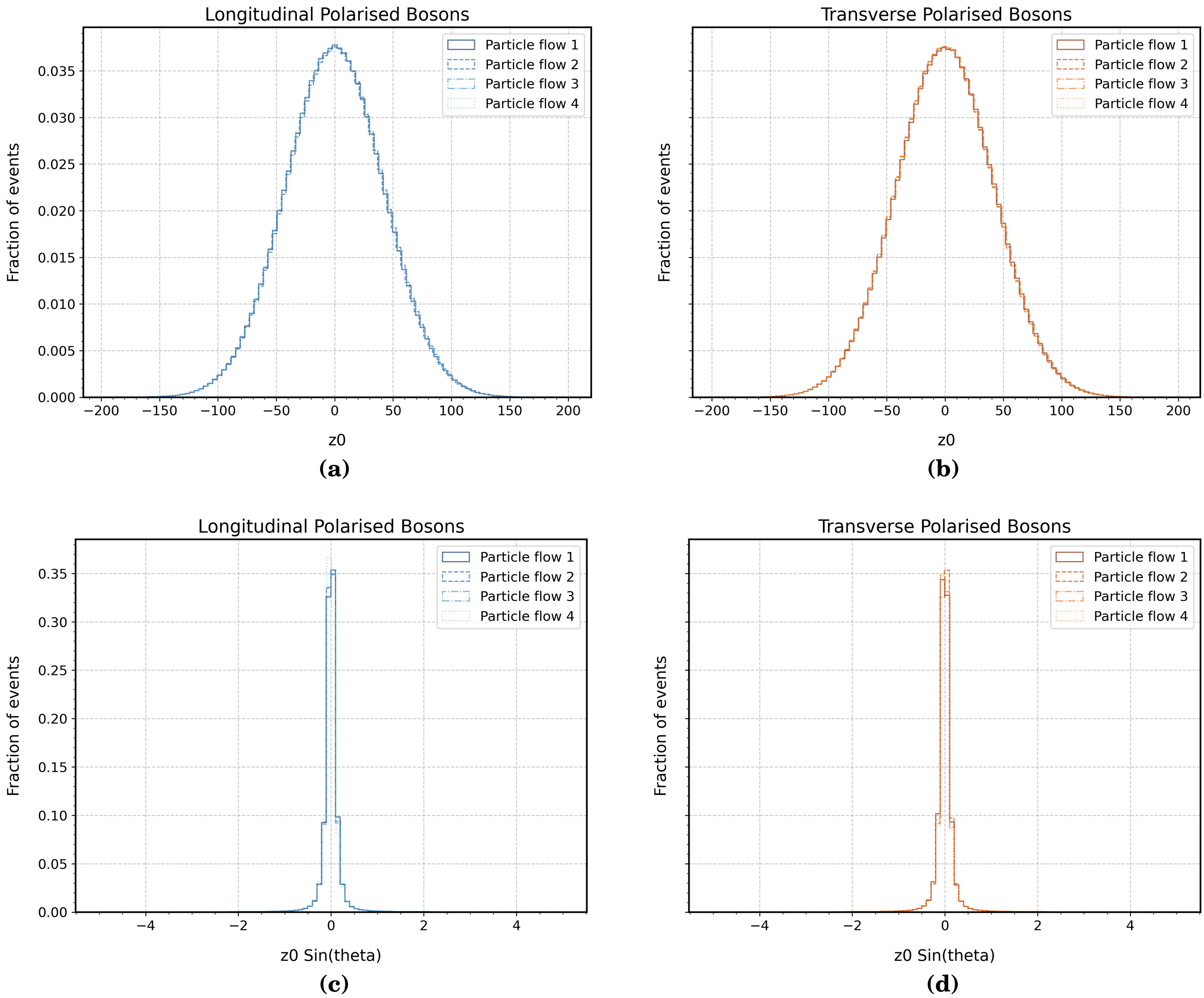


Figure 58: Distributions of some extra particle flow features for polarisation tagging, including the z_0 relative to the beamspot and $z_0 \sin\theta$, which again show minimal discrepancies.

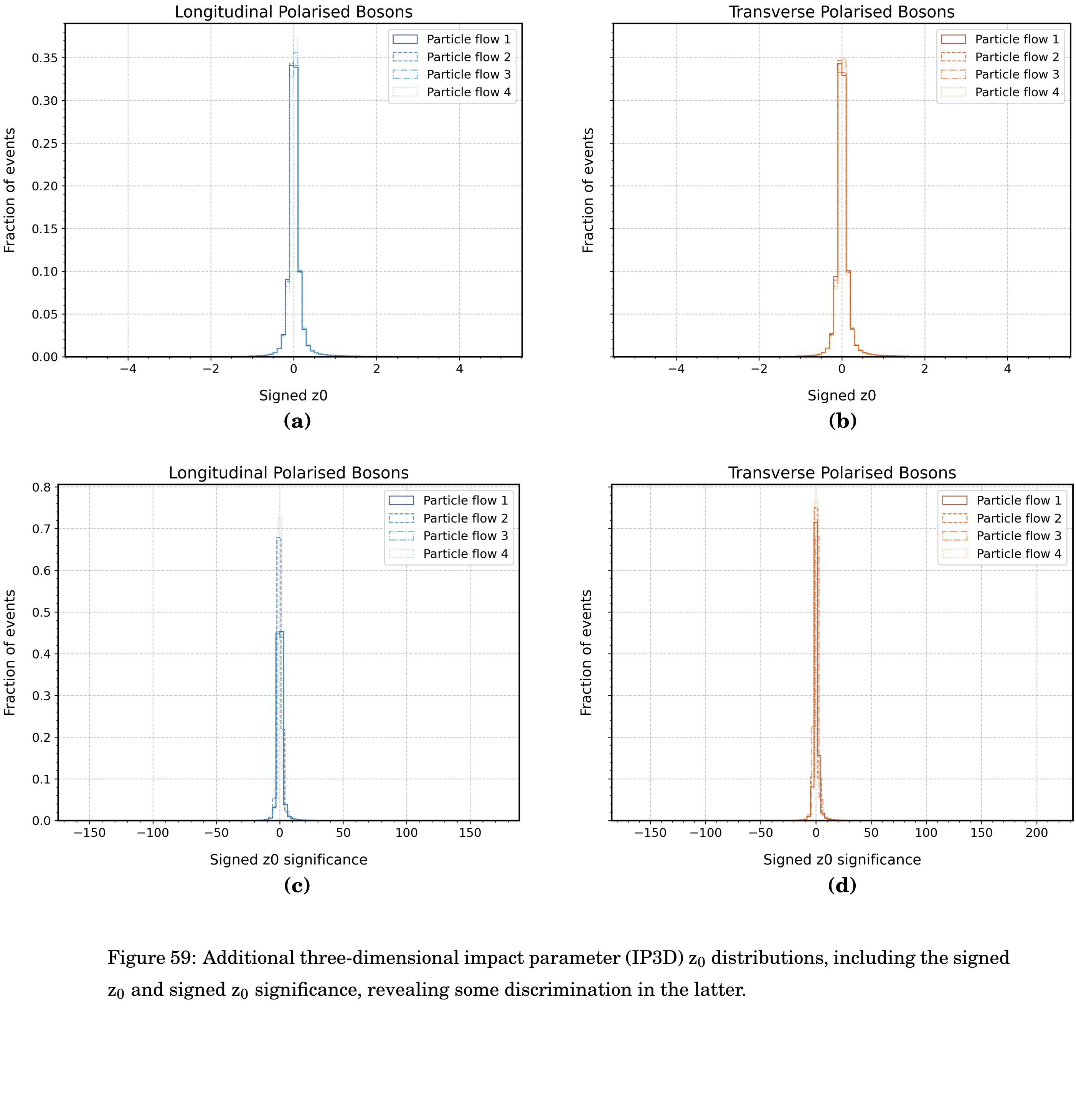


Figure 59: Additional three-dimensional impact parameter (IP3D) z_0 distributions, including the signed z_0 and signed z_0 significance, revealing some discrimination in the latter.

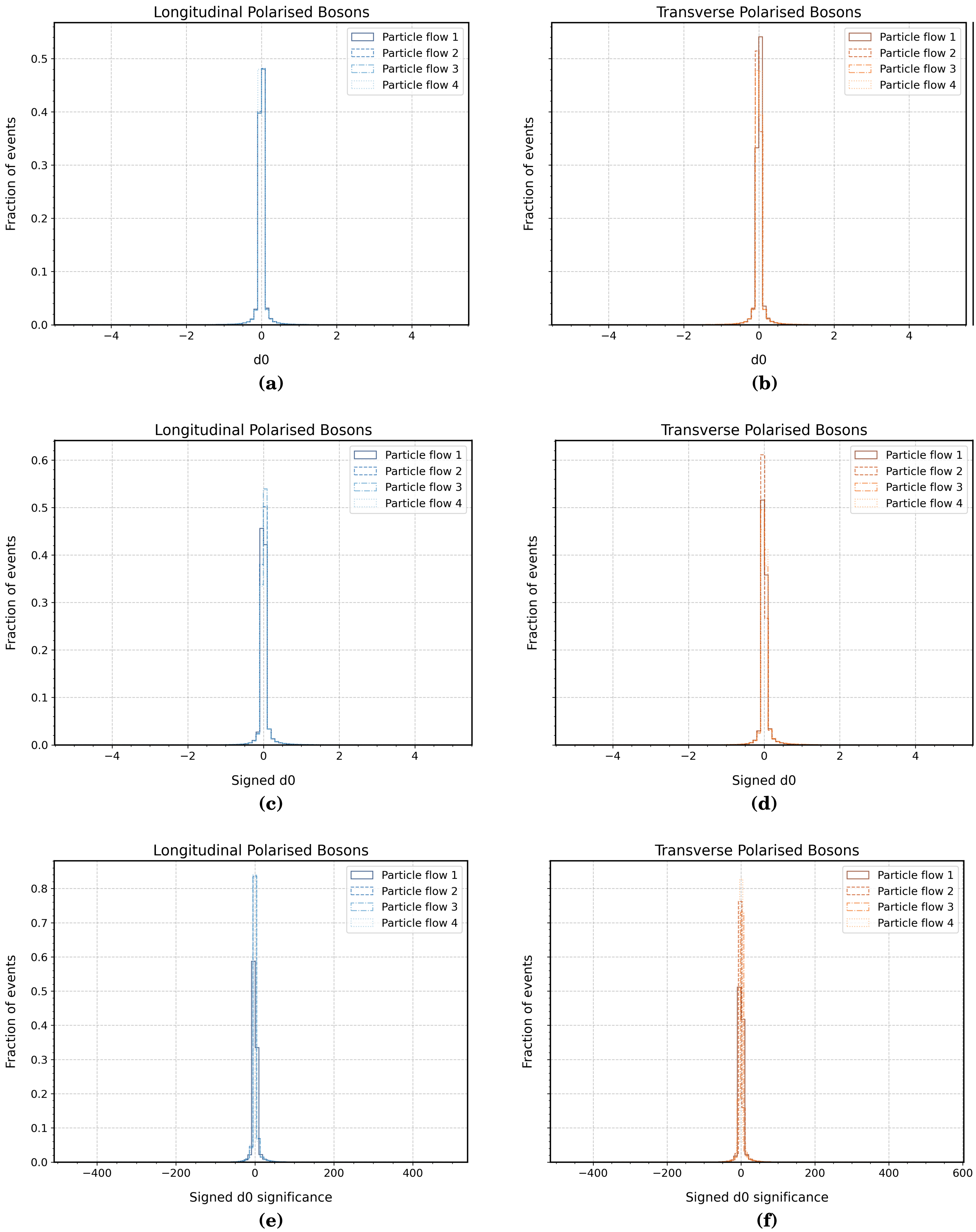


Figure 60: Distributions of three-dimensional impact parameter (IP3D) variables based on d_0, including the d_0, signed d_0 and signed d_0 significance. Some polarisation separation is observed.

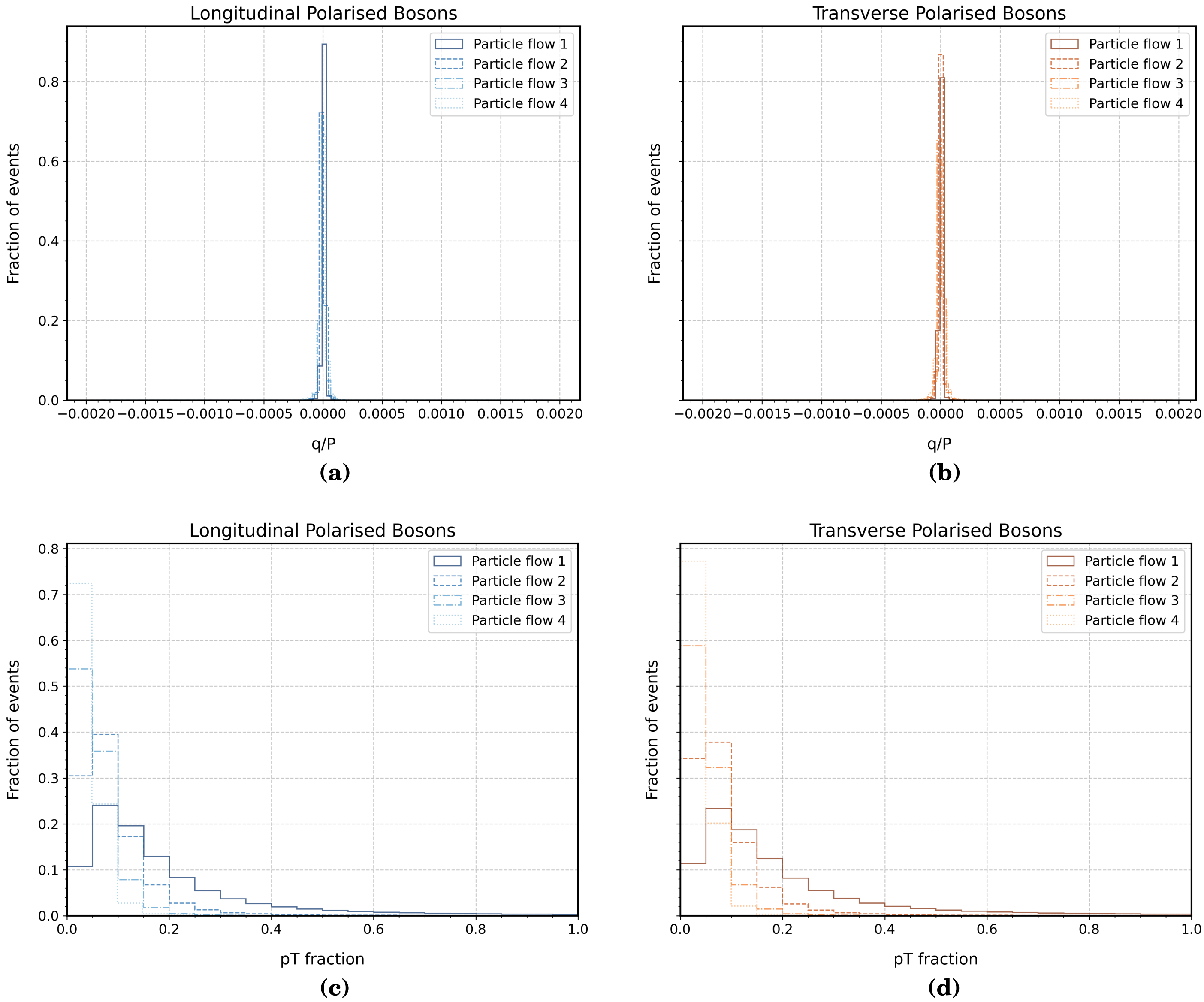


Figure 61: Distributions of additional particle flow features for polarisation tagging, including q/P from the Perigee coordinate system and the transverse momentum fraction, showing limited discrimination power.

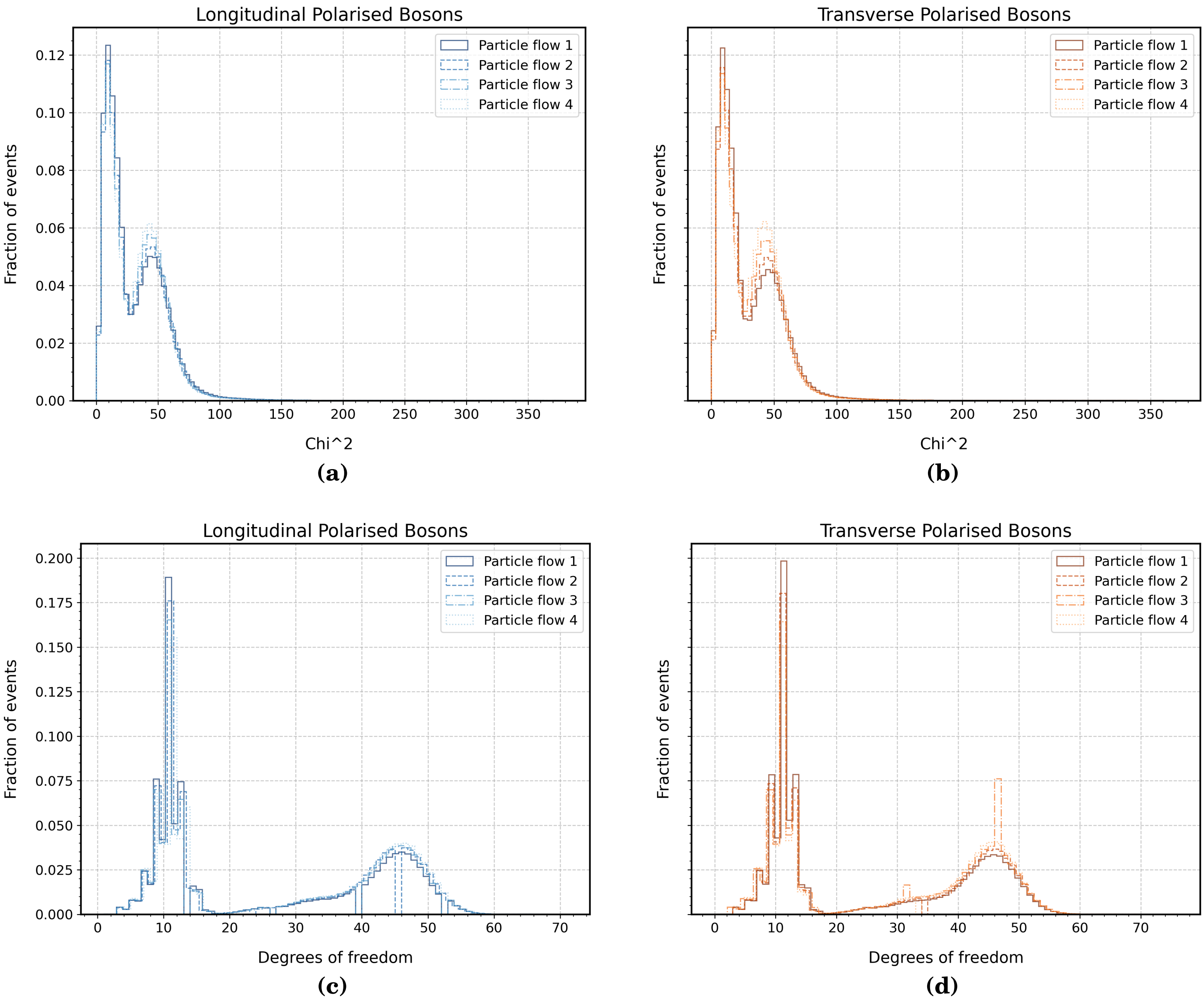


Figure 62: Distributions of additional non-discriminative particle flow features, including the χ^2 and numder of degrees of freedom.

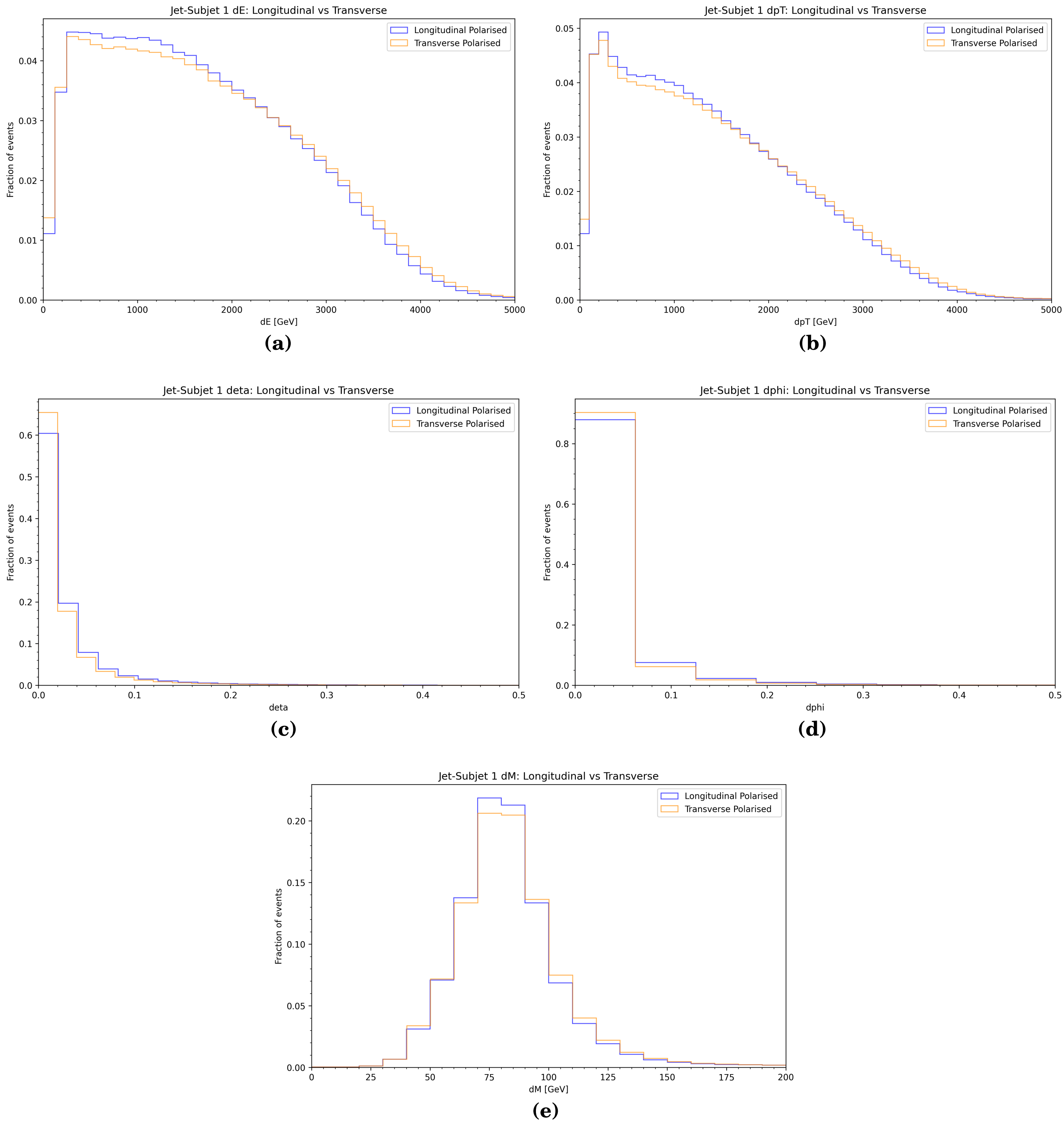
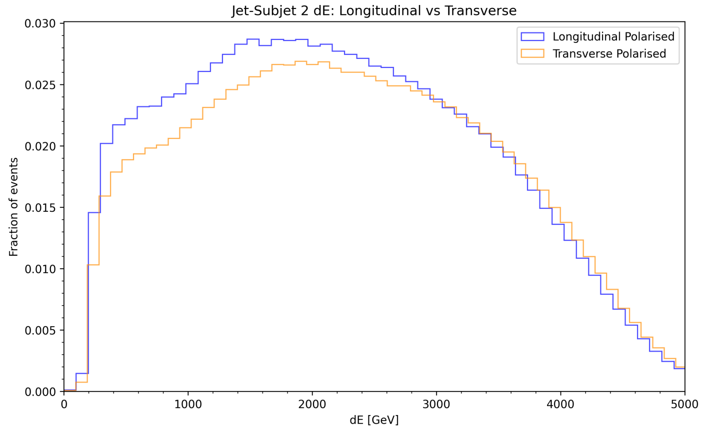
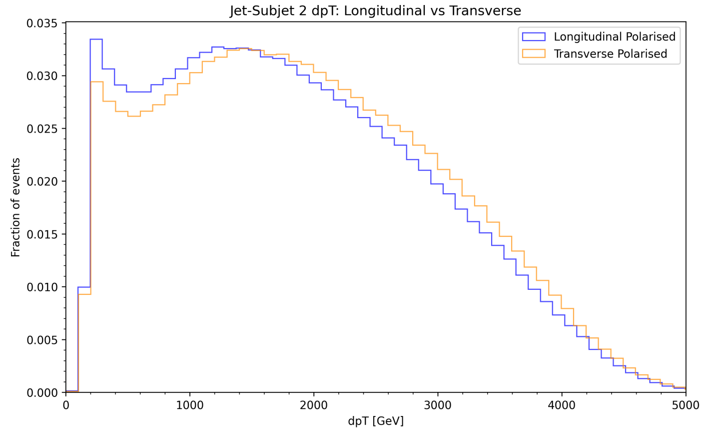


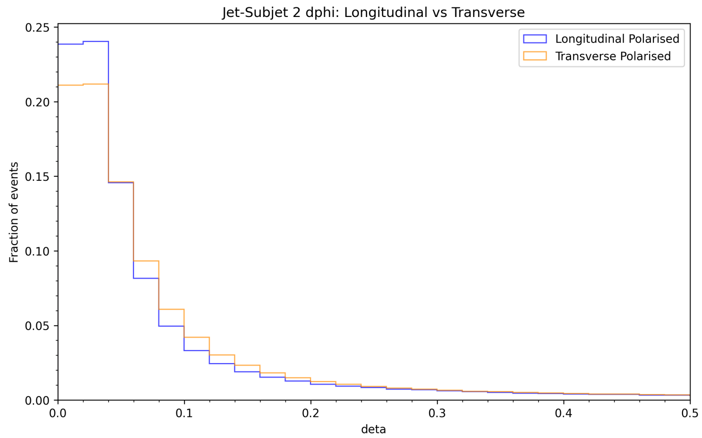
Figure 63: Distributions of engineered feature differences for the jet and the primary subjet in the context of polarisation tagging, including energy, p_T, η, ϕ, and mass. The distributions reveal minimal discrimination between the classes, indicating a need for further refinement in feature selection.



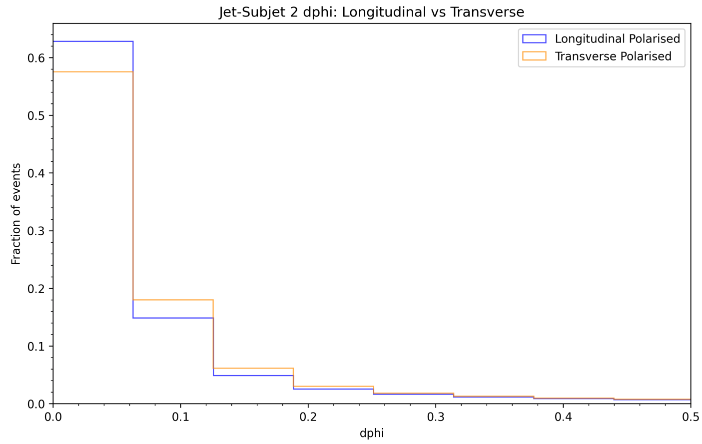
(a)



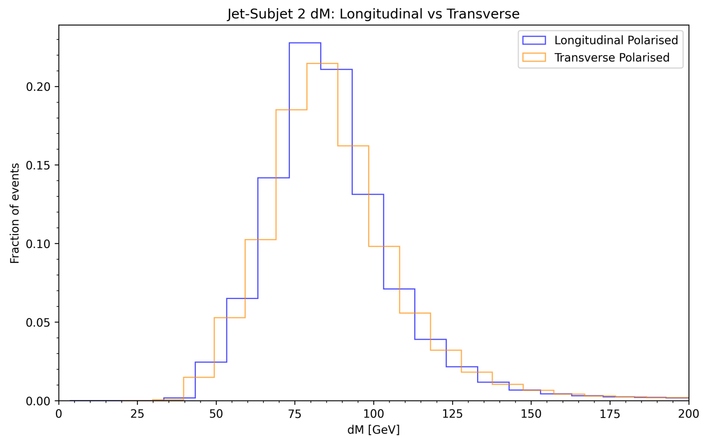
(b)



(c)



(d)



(e)

Figure 64: Distributions of engineered feature differences for the jet and the secondary subjet in the context of polarisation tagging, including energy, p_T, η, ϕ, and mass. While the distributions exhibit minimal discrimination between the classes, they show slightly better separation compared to Fig. 63.

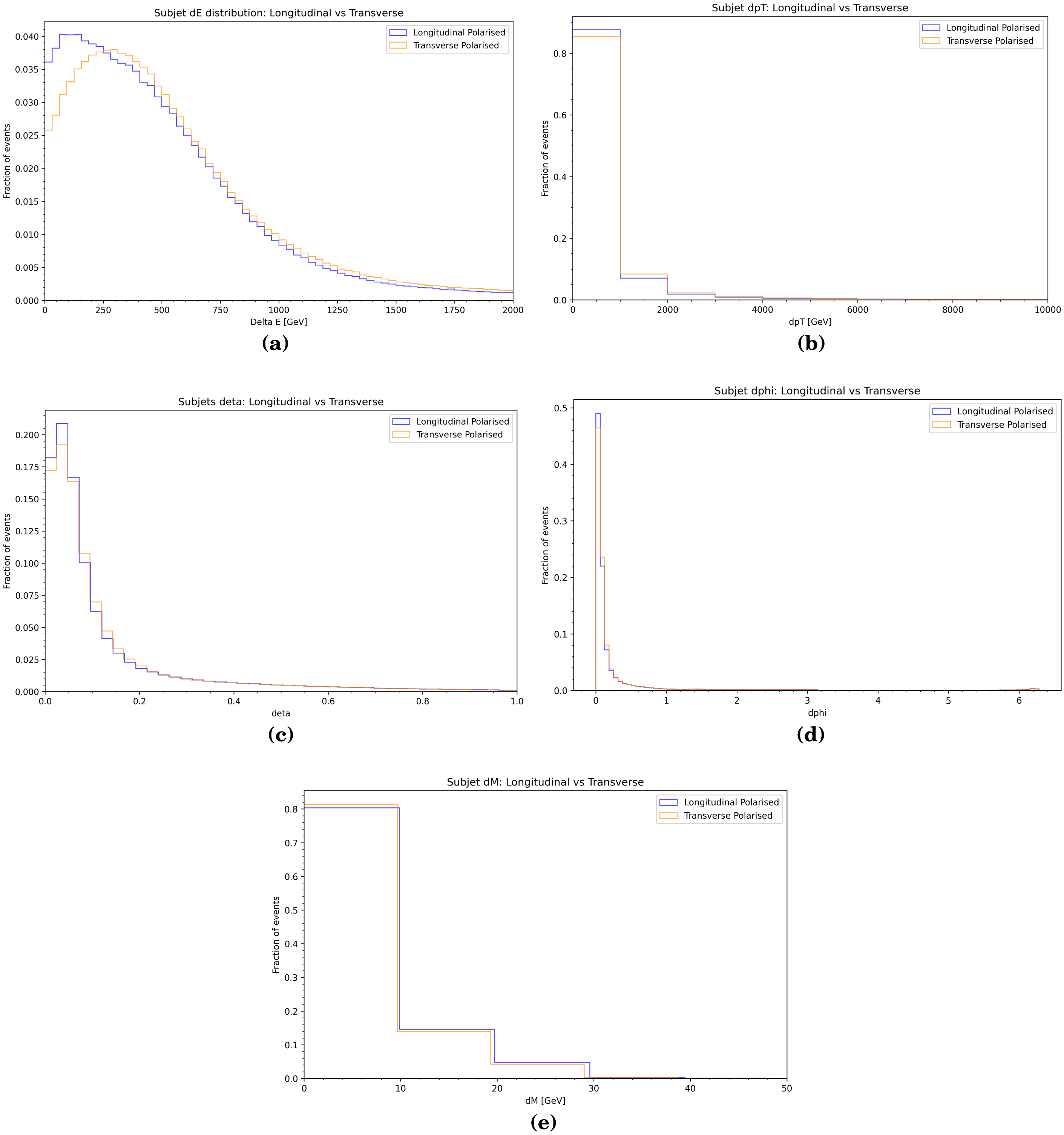


Figure 65: Distributions of engineered feature differences between the primary and secondary subjets in the context of polarisation tagging encompass energy, p_T, η, ϕ, and mass. The ΔE distribution demonstrates some expected discrimination, but the level is insufficient to make a significant impact and highlights the necessity for further refinement in feature selection.

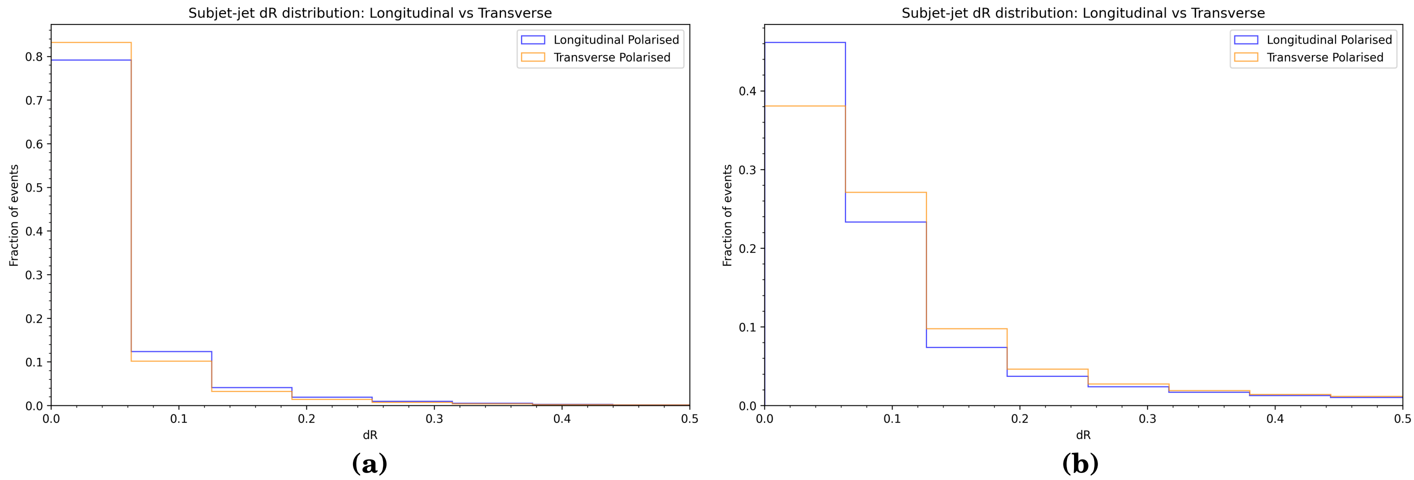


Figure 66: Distributions of the engineered angular distance features between the jet and subjets in the context of polarisation tagging. This ΔR is calculated using the angular differences between the jets as $\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. Panel (a) displays the angular distance for the primary subjet, while panel (b) illustrates the angular distance for the secondary subjet, which notably exhibits some discrimination between the two polarisation states.

The final part of this appendix contains the additional machine learning plots, including additional classification and regression performance metrics. Figure 67 compares the F1 score as a function of the classification threshold, reflecting their ability to balance precision and recall. This plot highlights the difficulty of polarisation classification for the DNN. Additionally, Figure 68 compares the AMS (Approximate Median Significance) and weighted accuracy. The BDT exhibits a more stable peak in AMS around a threshold of 0.4, whereas the DNN again starts at its peak and falls down directly. Finally, Figure 69 presents the trade-offs in signal and background classification accuracy as the threshold is adjusted. While the BDT shows a gradual decline in signal accuracy, the DNN demonstrates a more rapid decline, with background accuracy rising fast, indicating less stability and separation in classification performance.

The last plots (Figs. 70, 71, 72) illustrate performance outcomes for additional $\cos\theta$ regression approaches. The first plot displays the model's performance when regressing the full $\cos\theta$ range instead of $|\cos\theta|$. This method demonstrates notably poor performance, and this thesis did not successfully address this issue or improve predictions within this specific range. However, to distinguish between left- and right-handed polarisations, Figure 71 shows the results of regressing $\cos\theta$ while limiting the range to either positive or negative values. This approach yields performance similar to that of $|\cos\theta|$ regression.

The final plot showcases the $|\cos\theta|$ regression using a different loss function known as LogCosh Loss. This function computes the logarithm of the hyperbolic cosine of the prediction error, behaving like root-mean-square loss for small errors and root-mean-absolute loss for larger errors, making it less sensitive to outliers compared to mean square loss. The initial strategy involved training the model with MSE to eliminate large outliers and regressing the results shown in Figure 30. Subsequently, the model would be retrained using the LogCosh loss function to address smaller deviations. However, this approach counterintuitively pushed all predicted values further away from the target, ultimately degrading the model's performance. The subsequent strategies, shown in Figure 72 involved training solely with the LogCosh loss function and evaluating the results when retraining this model with the MSE function.

In conclusion, the plots in this appendix provide additional context and clarity regarding the feature distributions and validation metrics for the polarisation related studies.

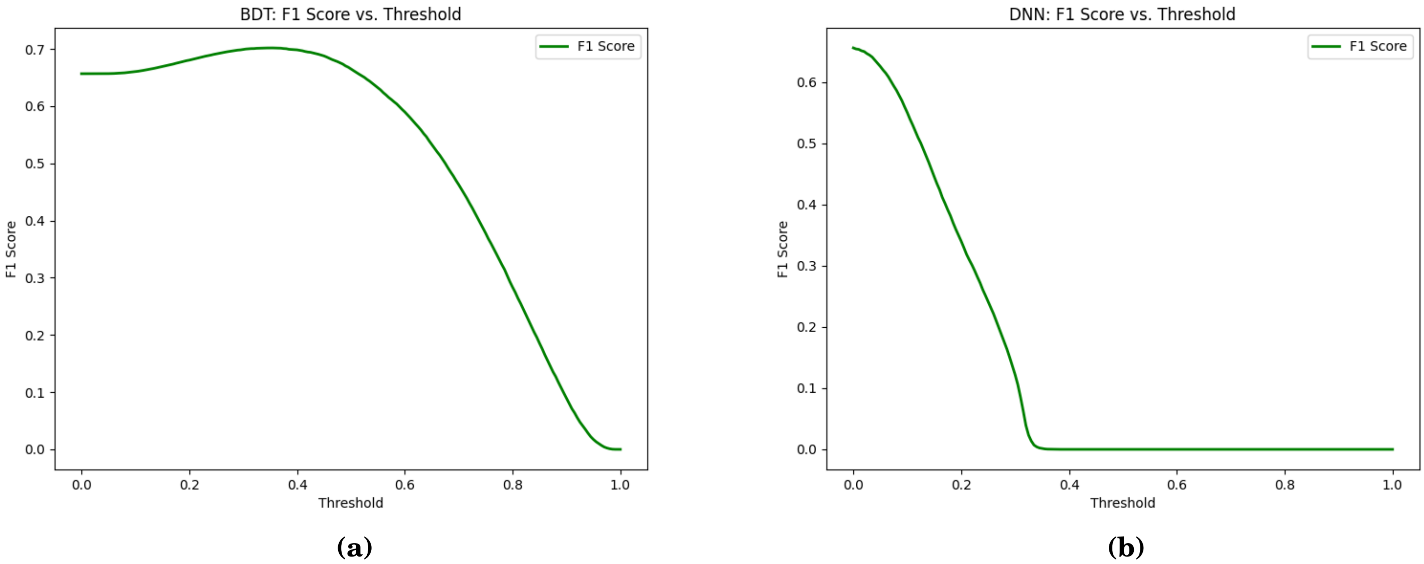
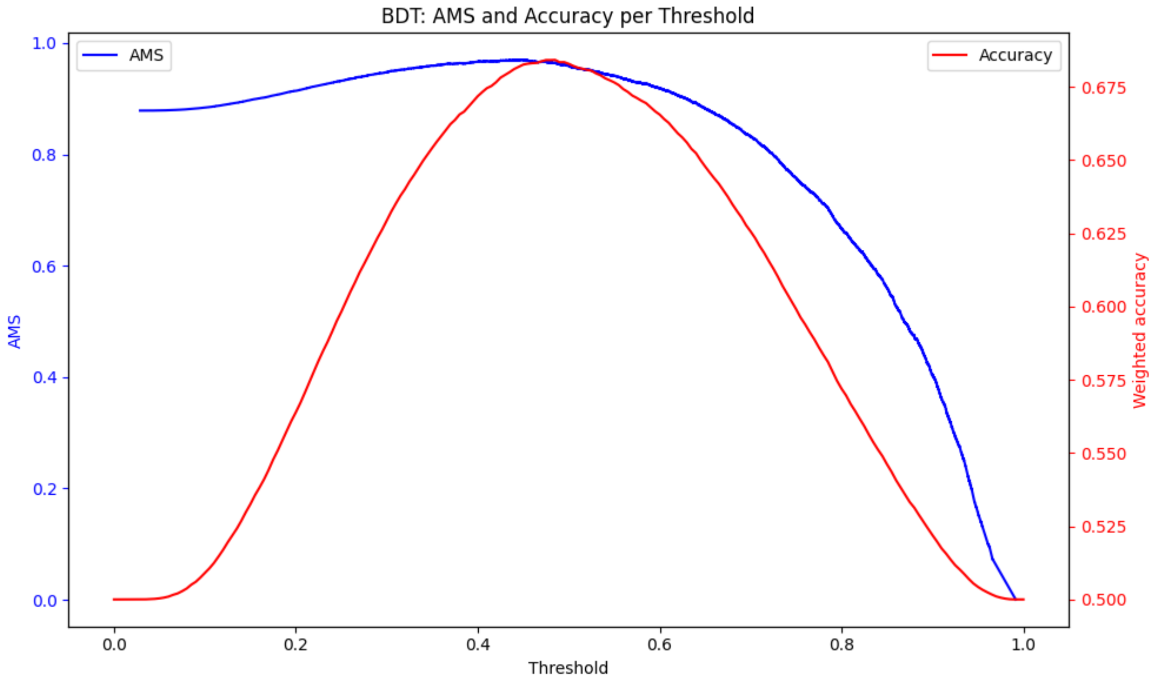
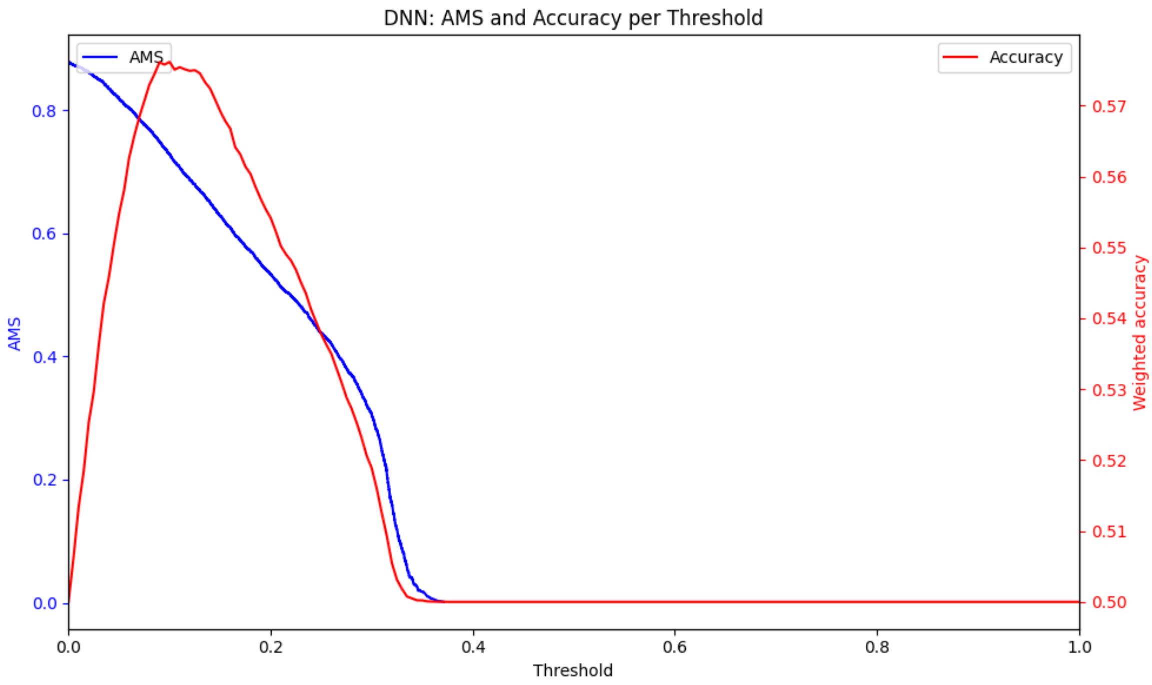


Figure 67: Comparison of F1 score versus threshold for the boosted decision tree (BDT) and deep neural network (DNN) models. The F1 score, a harmonic mean of precision and recall, measures the balance between correctly identifying signal and distinguishing it from background. In plot (a), the BDT starts with a relatively high F1 score around 0.65 at threshold zero, rising to a peak of approximately 0.7 around threshold 0.4, then gradually declining, approaching zero after threshold 0.8. Conversely, in panel (b), the DNN begins at an F1 score of 0.65, falls relatively sharp to zero, and remains flat from threshold 0.4. This suggests that the DNN struggles to classify polarisation effectively, with no predictions above a threshold of 0.4, highlighting its difficulty in distinguishing between polarisation states.

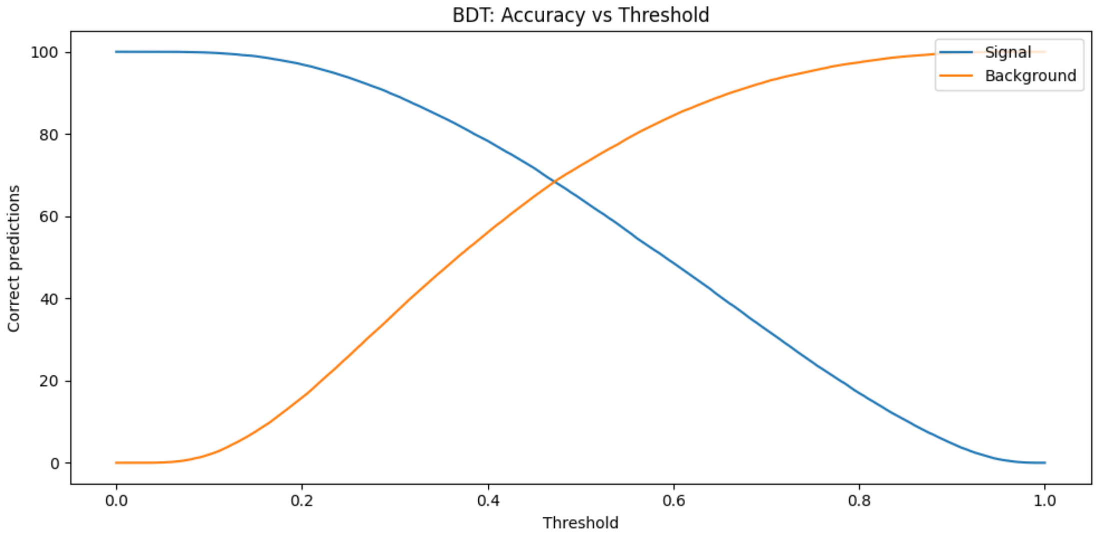


(a)

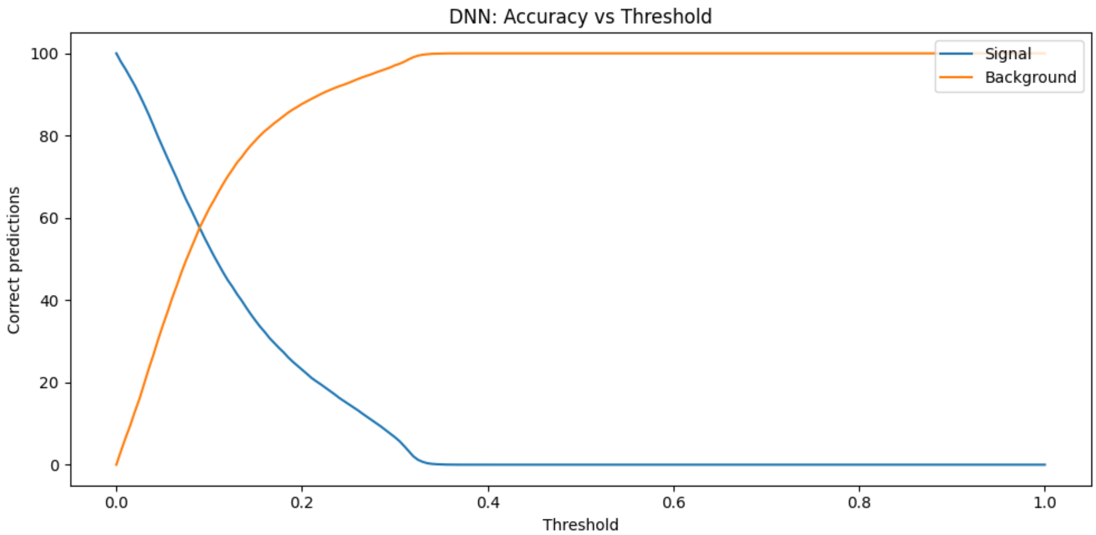


(b)

Figure 68: Comparison of performance metrics for the boosted decision tree (BDT) and deep neural network (DNN) models. The plots illustrate the AMS (blue curve) and weighted accuracy (red curve) as functions of the classification threshold. AMS (Approximate Median Significance) evaluates the statistical significance of a signal amidst background noise. For the BDT (a), the AMS peaks around a threshold of 0.4 before gradually decreasing, with a sharp drop beyond 0.8. The weighted accuracy, defined as $\sqrt{Acc^2_{\text{signal}} + Acc^2_{\text{bkg}}}$, forms a symmetric curve, reaching its maximum at a threshold of 0.5. In the DNN plot (b), the AMS starts high and steadily declines until a threshold of 0.3, after which it drops sharply to zero. The weighted accuracy peaks around a threshold of 0.2, then rapidly decreases to zero by 0.4, where it remains. These plots demonstrate the models' sensitivity to threshold selection, which influences both AMS and weighted accuracy.



(a)



(b)

Figure 69: Comparison of signal and background classification accuracy for the boosted decision tree (BDT) and deep neural network (DNN) models as a function of the classification threshold. In the upper plot (a), the BDT model shows signal accuracy (blue) starting high and progressively decreasing as the threshold increases, while background accuracy (orange) exhibits an almost perfect inverse relationship. The lower plot (b) for the DNN model shows a more pronounced shift toward lower thresholds, with signal accuracy rising more sharply and background accuracy declining more rapidly. These plots highlight the trade-off between identifying signal and background events as the classification threshold changes, revealing how each model balances these two aspects differently.

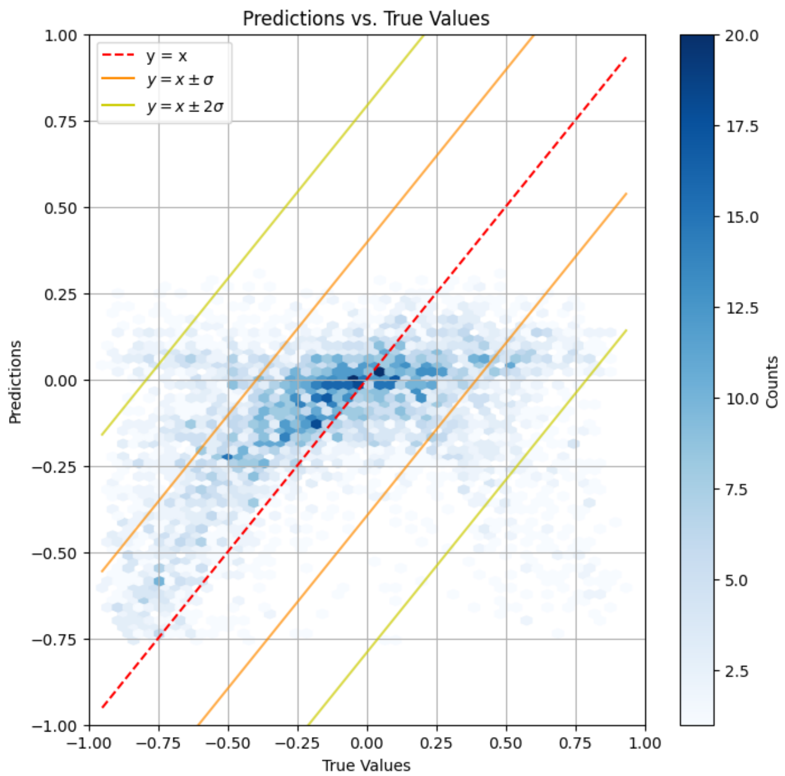
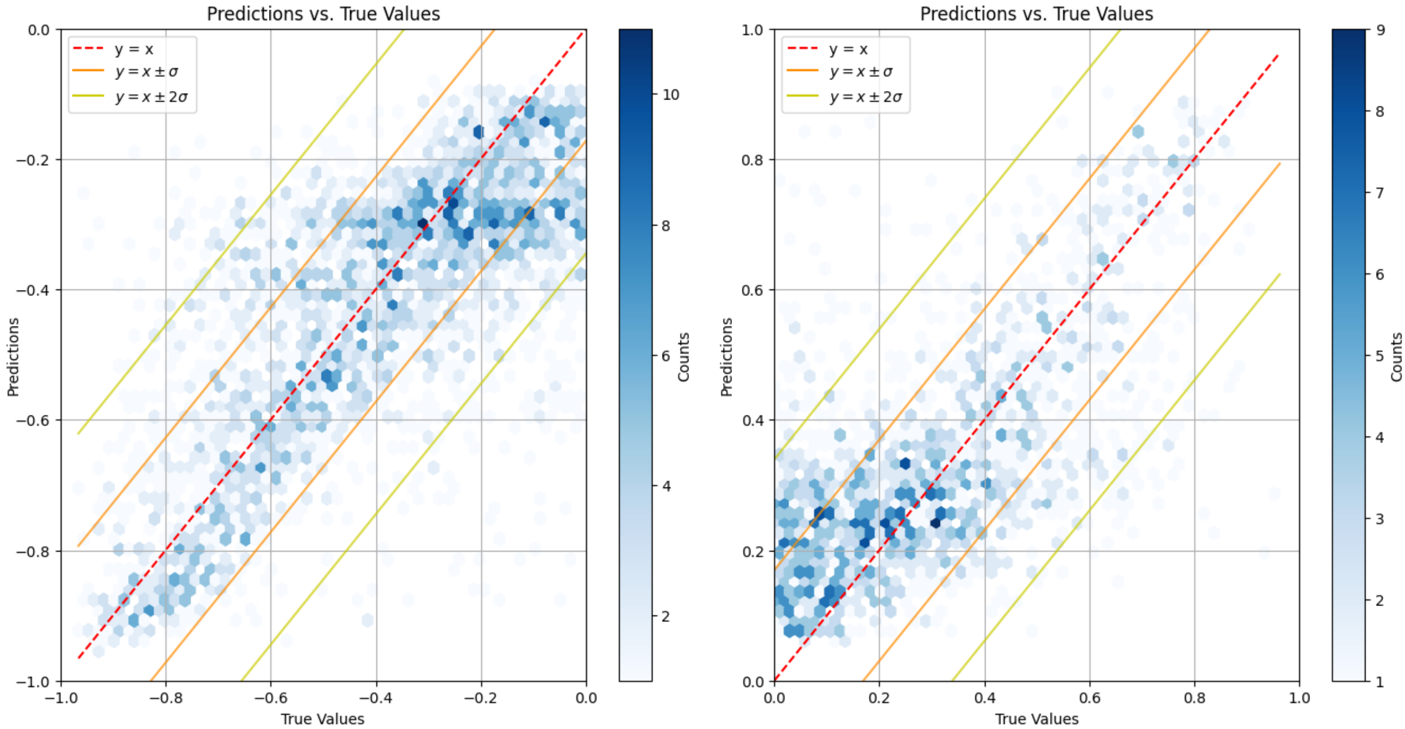
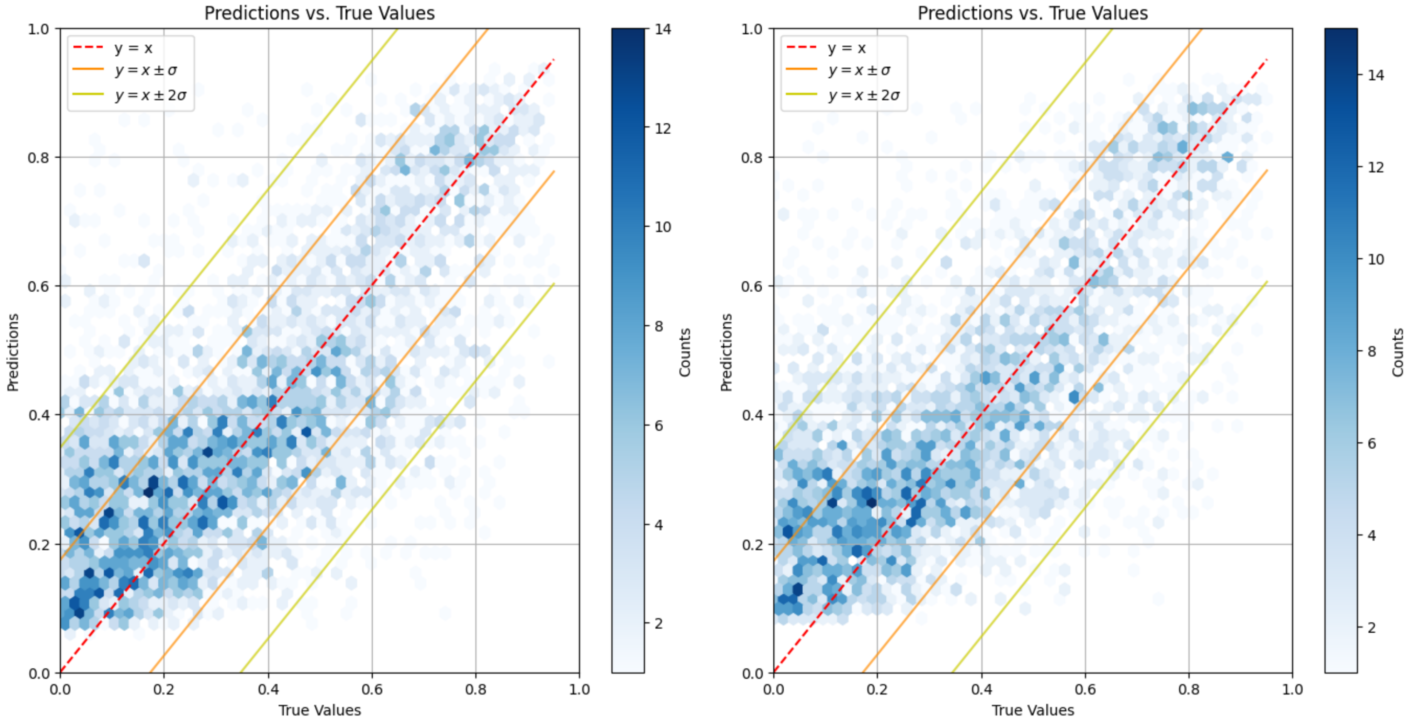


Figure 70: Predicted versus true values for the regression model output, encompassing both transverse and longitudinal polarisations. This plot demonstrates the model's performance in predicting $\cos\theta$ values for W and Z bosons. The diagonal red dashed line represents the ideal scenario where predictions align perfectly with true values (y=x). The yellow and orange lines indicate deviations of $\pm 1\sigma$ and $\pm 2\sigma$, respectively. The colour scale on the right illustrates the density of points, revealing areas of concentration for the model's predictions. The model exhibits poor performance, particularly for positive values, where it appears to have a prediction limit of 0.25. In contrast, predictions for lower values around -0.75 align more closely with the y=x line, although they still fall short of optimal accuracy.



(a) **(b)**

Figure 71: Predicted vs. true values for the regression model output, covering both transverse and longitudinal polarisations. The plot illustrates the model's performance for predicting $\cos\theta$ values of both W and Z bosons. Plot (a) displays the model's performance for negative $\cos\theta$ values, while plot (b) focuses on positive $\cos\theta$ values. The red dashed line represents perfect predictions (y=x), and the orange and yellow lines indicate $\pm 1\sigma$ and $\pm 2\sigma$ boundaries, respectively. The colour gradient represents prediction density, with darker regions indicating higher data concentrations. In both regimes the model performs reasonably well at larger $|\cos\theta|$ values, but accuracy declines near $\cos\theta \approx 0$, where predictions exhibit greater spread.



(a) **(b)**

Figure 72: Predicted vs. true values for the regression model output, comparing both transverse and longitudinal polarisations. The plot assesses the model's ability to predict $|\cos\theta|$ values for W and Z bosons. Plot (a) presents the results using the Logcosh loss function, while plot (b) shows the outcome after retraining the model with the MSE loss function. The red dashed line represents perfect predictions (y=x), with the orange and yellow lines marking $\pm 1\sigma$ and $\pm 2\sigma$ boundaries, respectively. The colour gradient indicates prediction density, with darker areas signifying higher concentrations of data points. Retraining slightly improves the model, especially for larger $|\cos\theta|$ values. However, both cases underperform compared to training exclusively with the MSE loss function.